Undergraduate Texts in Mathematics

Abbott: Understanding Analysis.
Anglin: Mathematics: A Concise History and Philosophy.
Readings in Mathematics.
Anglin/Lambek: The Heritage of Thales.
Readings in Mathematics.
Apostol: Introduction to Analytic Number Theory. Second edition.
Armstrong: Basic Topology.
Armstrong: Groups and Symmetry.
Axler: Linear Algebra Done Right. Second edition.
Beardon: Limits: A New Approach to Real Analysis.
Bak/Newman: Complex Analysis. Second edition.
Banchoff/Wermer: Linear Algebra Through Geometry. Second edition.
Berberian: A First Course in Real Analysis.
Bix: Conics and Cubics: A Concrete Introduction to Algebraic Curves.
Brémaud: An Introduction to Probabilistic Modeling.
Bressoud: Factorization and Primality Testing.
Bressoud: Second Year Calculus.
Readings in Mathematics.
Brickman: Mathematical Introduction to Linear Programming and Game Theory.
Browder: Mathematical Analysis: An Introduction.
Buchmann: Introduction to Cryptography.
Buskes/van Rooij: Topological Spaces: From Distance to Neighborhood.
Callahan: The Geometry of Spacetime: An Introduction to Special and General Relavitity.
Carter/van Brunt: The Lebesgue–Stieltjes Integral: A Practical Introduction.
Cederberg: A Course in Modern Geometries. Second edition.
Childs: A Concrete Introduction to Higher Algebra. Second edition.
Chung: Elementary Probability Theory with Stochastic Processes. Third edition.
Cox/Little/O'Shea: Ideals, Varieties, and Algorithms. Second edition.
Croom: Basic Concepts of Algebraic Topology.
Curtis: Linear Algebra: An Introductory Approach. Fourth edition.
Devlin: The Joy of Sets: Fundamentals of Contemporary Set Theory. Second edition.
Dixmier: General Topology.
Driver: Why Math?
Ebbinghaus/Flum/Thomas: Mathematical Logic. Second edition.
Edgar: Measure, Topology, and Fractal Geometry.
Elaydi: An Introduction to Difference Equations. Second edition.
Exner: An Accompaniment to Higher Mathematics.
Exner: Inside Calculus.
Fine/Rosenberger: The Fundamental Theory of Algebra.
Fischer: Intermediate Real Analysis.
Flanigan/Kazdan: Calculus Two: Linear and Nonlinear Functions. Second edition.
Fleming: Functions of Several Variables. Second edition.
Foulds: Combinatorial Optimization for Undergraduates.
Foulds: Optimization Techniques: An Introduction.
Franklin: Methods of Mathematical Economics.
Frazier: An Introduction to Wavelets Through Linear Algebra.
Gamelin: Complex Analysis.
Gordon: Discrete Probability.
Hairer/Wanner: Analysis by Its History.
Readings in Mathematics.

(continued after index)

Charles Chapman Pugh

Real Mathematical Analysis

With 133 Illustrations

Springer

Charles Chapman Pugh
Mathematics Department
University of California at Berkeley
Berkeley, CA 94720-3840
USA

Editorial Board

S. Axler
Mathematics Department
San Francisco State University
San Francisco, CA 94132
USA

F.W. Gehring
Mathematics Department
East Hall
University of Michigan
Ann Arbor, MI 48109
USA

K.A. Ribet
Mathematics Department
University of California, Berkeley
Berkeley, CA 94720-3840
USA

Mathematics Subject Classification (2000): 26-01

Library of Congress Cataloging-in-Publication Data
Pugh, C.C. (Charles Chapman), 1940–
Real mathematical analysis/Charles Chapman Pugh.
p. cm. — (Undergraduate texts in mathematics)
Includes bibliographical references and index.
ISBN 0-387-95297-7 (alk. paper)
1. Mathematical analysis. I. Title. II. Series.
QA300.P994 2001
515—dc21 2001032814

Printed on acid-free paper.

© 2002 Springer-Verlag New York, Inc.
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Production managed by Michael Koy; manufacturing supervised by Erica Bresler.
Formatted from the author's LaTeX2e files by TEXniques, Inc., Cambridge, MA.
Printed and bound by Maple-Vail Book Manufacturing Group, York, PA.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-95297-7 SPIN 10838675

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

To the students who have encouraged me
—especially A.W., D.H., and M.B.

Preface

Was plane geometry your favorite math course in high school? Did you like proving theorems? Are you sick of memorizing integrals? If so, real analysis could be your cup of tea. In contrast to calculus and elementary algebra, it involves neither formula manipulation nor applications to other fields of science. None. It is pure mathematics, and I hope it appeals to you, the budding pure mathematician.

Berkeley, California, USA CHARLES CHAPMAN PUGH

Contents

1 Real Numbers **1**

1 Preliminaries 1
2 Cuts 10
3 Euclidean Space 21
4 Cardinality 28
5* Comparing Cardinalities 34
6* The Skeleton of Calculus 36
Exercises 40

2 A Taste of Topology **51**

1 Metric Space Concepts 51
2 Compactness 76
3 Connectedness 82
4 Coverings 88
5 Cantor Sets 95
6* Cantor Set Lore 99
7* Completion 108
Exercises 115

3 Functions of a Real Variable **139**
1 Differentiation 139
2 Riemann Integration 154
3 Series 179
Exercises 186

4 Function Spaces **201**
1 Uniform Convergence and $C^0[a, b]$ 201
2 Power Series 211
3 Compactness and Equicontinuity in C^0 213
4 Uniform Approximation in C^0 217
5 Contractions and ODE's 228
6* Analytic Functions 235
7* Nowhere Differentiable Continuous Functions 240
8* Spaces of Unbounded Functions 248
Exercises 251

5 Multivariable Calculus **267**
1 Linear Algebra 267
2 Derivatives 271
3 Higher derivatives 279
4 Smoothness Classes 284
5 Implicit and Inverse Functions 286
6* The Rank Theorem 290
7* Lagrange Multipliers 296
8 Multiple Integrals 300
9 Differential Forms 313
10 The General Stokes' Formula 325
11* The Brouwer Fixed Point Theorem 334
Appendix A: Perorations of Dieudonné 337
Appendix B: The History of Cavalieri's Principle 338
Appendix C: A Short Excursion into the Complex Field 339
Appendix D: Polar Form 340
Appendix E: Determinants 342
Exercises 345

6 Lebesgue Theory **363**
1 Outer measure 363
2 Measurability 367
3 Regularity 371
4 Lebesgue integrals 376
5 Lebesgue integrals as limits 383
6 Italian Measure Theory 387
7 Vitali coverings and density points 391
8 Lebesgue's Fundamental Theorem of Calculus 396
9 Lebesgue's Last Theorem 401
Appendix A: Translations and Nonmeasurable sets 407
Appendix B: The Banach-Tarski Paradox 409
Appendix C: Riemann integrals as undergraphs 409
Appendix D: Littlewood's Three Principles 411
Appendix E: Roundness 412
Appendix F: Money 413
Suggested Reading 414
Bibliography 415
Exercises 417

Index **431**

1
Real Numbers

1 Preliminaries

Before discussing the system of real numbers it is best to make a few general remarks about mathematical outlook.

Language

By and large, mathematics is expressed in the language of set theory. Your first order of business is to get familiar with its vocabulary and grammar. A set is a collection of elements. The elements are members of the set and are said to belong to the set. For example, $\mathbb{N}$ denotes the set of **natural numbers**, $1, 2, 3, \ldots$. The members of $\mathbb{N}$ are whole numbers greater than or equal to 1. Is 10 a member of $\mathbb{N}$? Yes, 10 belongs to $\mathbb{N}$. Is π a member of $\mathbb{N}$? No. Is 0 a member of $\mathbb{N}$? No. We write

$$x \in A \quad \text{and} \quad y \notin B$$

to indicate that the element x is a member of the set A and y is not a member of B. Thus, $6819 \in \mathbb{N}$ and $\sqrt{2} \notin \mathbb{N}$.

We try to write capital letters for sets and small letters for elements of sets. Other standard sets have standard names. The set of **integers** is denoted by $\mathbb{Z}$, which stands for the German word *zahlen*. (An integer is a positive

whole number, zero, or a negative whole number.) Is $\sqrt{2} \in \mathbb{Z}$? No, $\sqrt{2} \notin \mathbb{Z}$. How about -15? Yes, $-15 \in \mathbb{Z}$.

The set of **rational numbers** is called $\mathbb{Q}$, which stands for "quotient." (A rational number is a fraction of integers, the denominator being nonzero.) Is $\sqrt{2}$ a member of $\mathbb{Q}$? No, $\sqrt{2}$ does not belong to $\mathbb{Q}$. Is π a member of $\mathbb{Q}$? No. Is 1.414 a member of $\mathbb{Q}$? Yes.

You should practice reading the notation "$\{x \in A$:" as "the set of x that belong to A such that." The **empty set** is the collection of no elements and is denoted by $\emptyset$. Is 0 a member of the empty set? No, $0 \notin \emptyset$.

A **singleton set** has exactly one member. It is denoted as $\{x\}$ where x is the member. Similarly if exactly two elements x and y belong to a set, the set is denoted as $\{x, y\}$.

If A and B are sets and each member of A also belongs to B then A is a subset of B and A is contained in B. We write[†]

$$A \subset B.$$

Is $\mathbb{N}$ a subset of $\mathbb{Z}$? Yes. Is it a subset of $\mathbb{Q}$? Yes. If A is a subset of B and B is a subset of C, does it follow that A is a subset of C? Yes. Is the empty set a subset of $\mathbb{N}$? Yes, $\emptyset \subset \mathbb{N}$. Is 1 a subset of $\mathbb{N}$? No, but the singleton set $\{1\}$ is a subset of $\mathbb{N}$. Two sets are equal if each member of one belongs to the other. Each is a subset of the other. This is how you prove two sets are equal: show that each element of the first belongs to the second, and each element of the second belongs to the first.

The union of the sets A and B is the set $A \cup B$, each of whose elements belongs to either A, or to B, or to both A and to B. The intersection of A and B is the set $A \cap B$ each of whose elements belongs to both A and to B. If $A \cap B$ is the empty set then A and B are **disjoint**. The **symmetric difference** of A and B is the set $A \Delta B$ each of whose elements belongs to A but not to B, or belongs to B but not to A. The **difference** of A to B is the set $A \setminus B$ whose elements belong to A but not to B. See Figure 1.

A **class** is a collection of sets. The sets are members of the class. For example we could consider the class $\mathcal{E}$ of sets of even natural numbers. Is the set $\{2, 15\}$ a member of $\mathcal{E}$? No. How about the singleton set $\{6\}$? Yes. How about the empty set? Yes, each element of the empty set is even.

When is one class a subclass of another? When each member of the former belongs also to the latter. For example the class $\mathcal{T}$ of sets of positive integers divisible by 10 is a subclass of $\mathcal{E}$, the class of sets of even natural

[†] When some mathematicians write $A \subset B$ they mean that A is a subset of B, but $A \neq B$. We do *not* adopt this convention. We accept $A \subset A$.

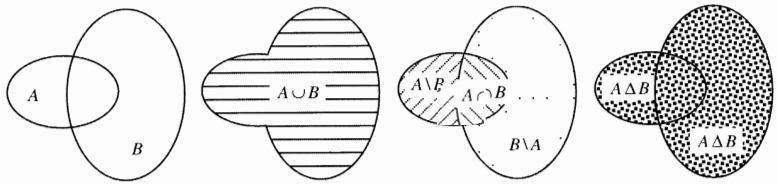


Figure 1 Venn diagrams of union, intersection, and differences.

numbers, and we write $\mathcal{T} \subset \mathcal{E}$. Each set that belongs to the class $\mathcal{T}$ also belongs to the class $\mathcal{E}$. Consider another example. Let $\mathcal{S}$ be the class of singleton subsets of $\mathbb{N}$ and $\mathcal{D}$ be the class of subsets of $\mathbb{N}$ each of which has exactly two elements. Thus $\{10\} \in \mathcal{S}$ and $\{2, 6\} \in \mathcal{D}$. Is $\mathcal{S}$ a subclass of $\mathcal{D}$? No. The members of $\mathcal{S}$ are singleton sets and they are not members of $\mathcal{D}$. Rather they are subsets of members of $\mathcal{D}$. Note the distinction, and think about it.

Here is an analogy. Each citizen is a member of his or her country – I am an element of the USA and Tony Blair is an element of the UK. Each country is a member of the United Nations. Are citizens members of the UN? No, countries are members of the UN.

In the same vein is the concept of an **equivalence relation** on a set S. It is a relation $s \sim s'$ that holds between some members $s, s' \in S$ and it satisfies three properties: For all $s, s', s'' \in S$

(a) $s \sim s$.
(b) $s \sim s'$ implies that $s' \sim s$.
(c) $s \sim s' \sim s''$ implies that $s \sim s''$.

The equivalence relation breaks S into disjoint subsets called **equivalence classes**[†] defined by mutual equivalence: the equivalence class containing s consists of all elements $s' \in S$ equivalent to s and is denoted $[s]$. The element s is a **representative** of its equivalence class. See Figure 2. Think again of citizens and countries. Say two citizens are equivalent if they are citizens of the same country. The world of equivalence relations is egalitarian: I represent my equivalence class USA just a s much as does the President.

Truth

When is a mathematical statement accepted as true? Generally, mathematicians would answer "Only when it has a proof inside a familiar mathematical

[†] The phrase "equivalence class" is standard and widespread, although it would be more consistent with the idea that a class is a collection of sets to refer instead to an "equivalence set."

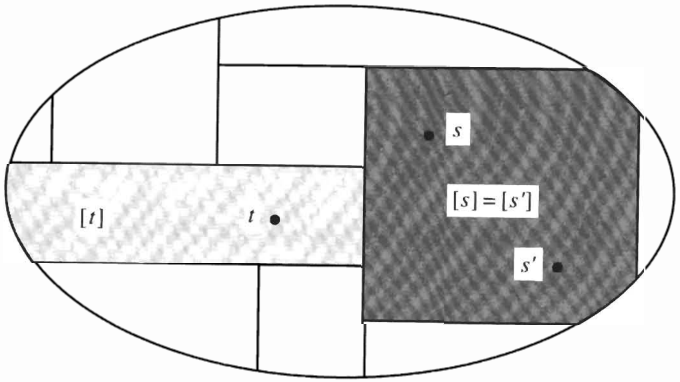


Figure 2 Equivalence classes and representatives.

framework." A picture may be vital in getting you to believe a statement. An analogy with something you know to be true may help you understand it. An authoritative teacher may force you to parrot it. A formal proof, however, is the ultimate and only reason to accept a mathematical statement as true. A recent debate in Berkeley focused the issue for me. According to a math teacher from one of our local private high schools, his students found proofs in mathematics were of little value, especially compared to "convincing arguments." Besides, the mathematical statements were often seen as obviously true and in no need of formal proof anyway. I offer you a paraphrase of Bob Osserman's response.

> But a convincing argument is not a proof. A mathematician generally wants both, and certainly would be less likely to accept a convincing argument by itself than a formal proof by itself. Least of all would a mathematician accept the proposal that we should generally replace proofs with convincing arguments.
>
> There has been a tendency in recent years to take the notion of proof down from its pedestal. Critics point out that standards of rigor change from century to century. New gray areas appear all the time. Is a proof by computer an acceptable proof? Is a proof that is spread over many journals and thousands of pages, that is too long for any one person to master, a proof? And of course, venerable Euclid is full of flaws, some filled in by Hilbert, others possibly still lurking.

> Clearly it is worth examining closely and critically the most basic notion of mathematics, that of proof. On the other hand, it is important to bear in mind that all distinctions and niceties about what precisely constitutes a proof are mere quibbles compared to the enormous gap between any generally accepted version of a proof and the notion of a convincing argument. Compare Euclid, with all his flaws to the most eminent of the ancient exponents of the convincing argument — Aristotle. Much of Aristotle's reasoning was brilliant, and he certainly convinced most thoughtful people for over a thousand years. In some cases his analyses were exactly right, but in others, such as heavy objects falling faster than light ones, they turned out to be totally wrong. In contrast, there is not to my knowledge a single theorem stated in Euclid's *Elements* that in the course of two thousand years turned out to be false. That is quite an astonishing record, and an extraordinary validation of proof over convincing argument.

Here are some guidelines for writing a rigorous mathematical proof. See also Exercise 0.

1. Name each object that appears in your proof. (For instance, you might begin your proof with a phrase, "consider a set X, and elements x, y that belong to X," etc.)
2. Draw a diagram that captures how these objects relate, and extract logical statements from it. Quantifiers precede the objects quantified; see below.
3. Proceed step-by-step, each step depending on the hypotheses, previously proved theorems, or previous steps in your proof.
4. Check for "rigor": all cases have been considered, all details have been tied down, and circular reasoning has been avoided.
5. Before you sign off on the proof, check for counter-examples and any implicit assumptions you made that could invalidate your reasoning.

Logic

Among the most frequently used logical symbols in math are the quantifiers $\forall$ and $\exists$. Read them always as "for each" and "there exists." Avoid reading $\forall$ as "for all," which in English has a more inclusive connotation. Another common symbol is $\Rightarrow$. Read it as "implies."

The rules of correct mathematical grammar are simple: quantifiers appear at the beginning of a sentence, they modify only what follows them in the sentence, and assertions occur at the end of the sentence. Here is an example.

(1)
For each integer n there is a prime number p which is greater than n.

In symbols the sentence reads

$$\forall n \in \mathbb{Z} \quad \exists p \in P \quad \text{such that} \quad p > n,$$

where P denotes the set of prime numbers. (A **prime number** is a whole number greater than 1 whose only divisors in $\mathbb{N}$ are itself and 1.) In English, the same idea can be re-expressed as

(2) *Every integer is less than some prime number*

or

(3) *A prime number can always be found*
which is greater than any given integer.

These sentences are correct in English grammar, but disastrously WRONG when transcribed directly into mathematical grammar. They translate into disgusting mathematical gibberish:

(WRONG 2) $\forall n \in \mathbb{Z} \quad n < p \quad \exists p \in P$
(WRONG 3) $\exists p \in P \quad p > n \quad \forall n \in \mathbb{Z}.$

Moral Quantifiers first and assertions last. In stating a theorem, try to apply the same principle. Write the hypothesis first and the conclusion second. See Exercise 0.

The order in which quantifiers appear is also important. Contrast the next two sentences in which we switch the position of two quantified phrases.

(4) $(\forall n \in \mathbb{N}) \quad (\forall m \in \mathbb{N}) \quad (\exists p \in P) \quad \text{such that} \quad (nm < p).$

(5) $(\forall n \in \mathbb{N}) \quad (\exists p \in P) \quad \text{such that} \quad (\forall m \in \mathbb{N}) \quad (nm < p).$

(4) is a true statement but (5) is false. A quantifier modifies the part of a sentence that follows it but not the part that precedes it. This is another reason never to end with a quantifier.

Moral Quantifier order is crucial.

There is a point at which English and mathematical meaning diverge. It concerns the word "or." In mathematics "a or b" always means "a or b or both a and b," while in English it can mean "a or b but not both a and b." For example, Patrick Henry certainly would not have accepted both liberty and death in response to his cry of "Give me liberty or give me death." In mathematics, however, the sentence "17 is a prime or 23 is a prime" is correct even though both 17 and 23 are prime. Similarly, in mathematics a $\Rightarrow b$ means that if a is true then b is true but that b might also be true for reasons entirely unrelated to the truth of a. In English, $a \Rightarrow b$ is often confused with $b \Rightarrow a$.

Moral In mathematics, "or" is inclusive. It means *and/or*. In mathematics, $a \Rightarrow b$ is not the same as $b \Rightarrow a$.

It is often useful to form the negation or logical opposite of a mathematical sentence. The symbol $\sim$ is usually used for negation, despite the fact that the same symbol also indicates an equivalence relation. Mathematicians refer to this as an **abuse of notation**. Fighting a losing battle against abuse of notation, we write $\neg$ for negation. For example, if $m, n \in \mathbb{N}$ then $\neg(m < n)$ means it is not true that m is less than n. In other words

$$\neg(m < n) \quad \equiv \quad m \geq n.$$

(We use the symbol $\equiv$ to indicate that the two statements are equivalent.) Similarly, $\neg(x \in A)$ means it is not true that x belongs to A. In other words,

$$\neg(x \in A) \quad \equiv \quad x \notin A.$$

Double negation returns a statement to its original meaning. Slightly more interesting is the negation of "and" and "or." Just for now, let us use the symbols & for "and" and $\vee$ for "or." We claim

(6) $$\neg(a \ \& \ b) \quad \equiv \quad \neg a \vee \neg b.$$

(7) $$\neg(a \vee b) \quad \equiv \quad \neg a \ \& \ \neg b.$$

For if it is not the case that both a and b are true then at least one must be false. This proves (6), and (7) is similar. Implication also has such interpretations:

(8) $$a \Rightarrow b \quad \equiv \quad \neg a \Leftarrow \neg b \quad \equiv \quad \neg a \vee b.$$

(9) $$\neg(a \Rightarrow b) \quad \equiv \quad a \;\&\; \neg b.$$

What about the negation of a quantified sentence such as

$$\neg(\forall n \in \mathbb{N}, \exists p \in P \text{ such that } n < p).$$

The rule is: change each $\forall$ to $\exists$ and vice versa, leaving the order the same, and negate the assertion. In this case the negation is

$$\exists n \in \mathbb{N}, \quad \forall p \in P, \quad n \geq p.$$

In English it reads "There exists a natural number n, and for all primes p, $n \geq p$." The sentence has correct mathematical grammar but of course is false. To help translate from mathematics to readable English, a comma can be read as "and" or "such that."

All mathematical assertions take an implication form $a \Rightarrow b$. The hypothesis is a and the conclusion is b. If you are asked to prove $a \Rightarrow b$, there are several ways to proceed. First you may just see right away why a does imply b. Fine, if you are so lucky. Or you may be puzzled. Does a really imply b? Two routes are open to you. You may view the implication in its equivalent contrapositive form $\neg a \Leftarrow \neg b$ as in (8). Sometimes this will make things clearer. Or you may explore the possibility that a fails to imply b. If you can somehow deduce from the failure of a implying b a contradiction to a known fact (for instance if you can deduce the existence of a planar right triangle with legs x, y but $x^2 + y^2 \neq h^2$ where h is the hypotenuse) then you have succeeded in making an **argument by contradiction**. Clearly (9) is pertinent here. It tells you what it means that a fails to imply b, namely that a is true and, simultaneously, b is false.

Euclid's proof that $\mathbb{N}$ contains infinitely many prime numbers, is a classic example of this method. The hypothesis is that $\mathbb{N}$ is the set of natural numbers and that P is the set of prime numbers. The conclusion is that P is an infinite set. The proof of this fact begins with the phrase "Suppose not." It means: suppose, after all, that the set of prime numbers P is merely a finite set, and see where this leads you. It does not mean that we think P really is a finite set, and it is not a hypothesis of a theorem. Rather it just means that we will try to find out what awful consequences would follow from P being finite. In fact if P were[†] finite then it would consist of m

[†]In English grammar, the subjunctive mode indicates doubt, and I have written Euclid's proof in that form – "if P *were* finite" instead of "if P *is* finite," "each prime *would* divide N evenly," instead of "each prime *divides* N evenly," etc. At first it seems like a fine idea to write all arguments by contradiction in the subjunctive mode, exhibiting clearly their impermanence. Soon, however, the subjunctive and conditional language becomes ridiculously stilted and archaic. For consistency then, as much as possible, *use the present tense*.

numbers $p_1, \ldots, p_m$. Their product $N = 2 \cdot 3 \cdot 5 \cdot \cdots \cdot p_m$ would be evenly divisible (i.e., remainder 0 after division) by each p_i and therefore $N + 1$ would be evenly divisible by no prime (the remainder of p_i divided into $N + 1$ would always be 1), which would contradict the fact that every integer ≥ 2 can be factored as a product of primes. (The latter fact has nothing to do with P being finite or not.) Since the supposition that P is finite led to a contradiction of a known fact, prime factorization, the supposition was incorrect, and P is, after all, infinite.

Afficionados of logic will note our heavy use here of the "law of the excluded middle," to wit, that a mathematically meaningful statement is either true or false. The possibilities that it is neither true nor false, or that it is both true and false, are excluded.

Metaphor and Analogy

In high school English, you are taught that a metaphor is a figure of speech in which one idea or word is substituted for another to suggest a likeness or similarity. This can occur very simply as in "The ship plows the sea." Or it can be less direct, as in "his lawyers dropped the ball." What give a metaphor its power and pleasure are the secondary suggestions of similarity. Not only did the lawyers make a mistake, but it was their own fault, and, like an athlete who has dropped a ball, they could not follow through with their next legal action. A secondary implication is that their enterprise was just a game.

Often a metaphor associates something abstract to something concrete, as "Life is a journey." The preservation of inference from the concrete to the abstract in this metaphor suggests that like a journey, life has a beginning and an end, it progresses in one direction, it may have stops and detours, ups and downs, etc. The beauty of a metaphor is that hidden in a simple sentence like "Life is a journey" lurk a great many parallels, waiting to be uncovered by the thoughtful mind.

Metaphorical thinking pervades mathematics to a remarkable degree. It is often reflected in the language mathematicians choose to define new concepts. In his construction of the system of real numbers, Dedekind could have referred to $A|B$ as a "type-two, order preserving equivalence class," or worse, whereas "cut" is the right metaphor. It corresponds closely to one's physical intuition about the real line. See Figure 3. In his book, *Where Mathematics Comes From*, George Lakoff gives a comprehensive view of metaphor in mathematics.

An analogy is a shallow form of metaphor. It just asserts that two things are similar. Although simple, analogies can be a great help in accepting abstract concepts. When you travel from home to school, at first you are closer to home, and then you are closer to school. Somewhere there is a halfway stage in your journey. You *know* this, long before you study mathematics. So when a curve connects two points in a metric space (Chapter 2), you should expect that as a point "travels along the curve," somewhere it will be equidistant between the curve's endpoints. Reasoning by analogy is also referred to as "intuitive reasoning."

Moral Try to translate what you know of the real world to guess what is true in mathematics.

Two pieces of advice

A colleague of mine regularly gives his students an excellent piece of advice. When you confront a general problem and do not see how to solve it, make some extra hypotheses, and try to solve it then. If the problem is posed in n dimensions, try it first in two dimensions. If the problem assumes that some function is continuous, does it get easier for a differentiable function? The idea is to reduce an abstract problem to its simplest concrete manifestation, rather like a metaphor in reverse. At the minimum, look for at least one instance in which you can solve the problem, and build from there.

Moral If you do not see how to solve a problem in complete generality, first solve it in some special cases.

Here is the second piece of advice. Buy a notebook. In it keep a diary of your own opinions about the mathematics you are learning. Draw a picture to illustrate every definition, concept, and theorem.

2 Cuts

We begin at the beginning and discuss $\mathbb{R}$ = the system of all real numbers from a somewhat theological point of view. The current mathematics teaching trend treats the real number system $\mathbb{R}$ as a given — it is defined axiomatically. Ten or so of its properties are listed, called axioms of a complete ordered field, and the game becomes: deduce its other properties from the axioms. This is something of a fraud, considering that the entire structure of analysis is built on the real number system. For what if a system satisfying the axioms failed to exist? Then one would be studying the empty set! However, you need not take the existence of the real numbers on faith alone — we will give a concise mathematical proof of it.

It is reasonable to accept all grammar school arithmetic facts about

The set $\mathbb{N}$ of natural numbers, $1, 2, 3, 4, \ldots$.

The set $\mathbb{Z}$ of integers, $0, 1, -1, -2, 2, \ldots$.

The set $\mathbb{Q}$ of rational numbers p/q where p, q are integers, $q \neq 0$.

For example, we will admit without question facts like $2 + 2 = 4$, and laws like $a + b = b + a$ for rational numbers a, b. All facts you know about arithmetic involving integers or rational numbers are fair to use in homework exercises too.[†] It is clear that $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}$. Now $\mathbb{Z}$ improves $\mathbb{N}$ because it contains negatives and $\mathbb{Q}$ improves $\mathbb{Z}$ because it contains reciprocals. $\mathbb{Z}$ legalizes subtraction and $\mathbb{Q}$ legalizes division. Still, $\mathbb{Q}$ needs further improvement. It doesn't admit irrational roots such as $\sqrt{2}$ or transcendental numbers such as π. We aim to go a step beyond $\mathbb{Q}$, completing it to form $\mathbb{R}$ so that

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

As an example of the fact that $\mathbb{Q}$ is incomplete we have

1 Theorem *No number r in $\mathbb{Q}$ has square equal to 2; i.e., $\sqrt{2} \notin \mathbb{Q}$.*

Proof To prove that every $r = p/q$ has $r^2 \neq 2$ we show that $p^2 \neq 2q^2$. It is fair to assume that p and q have no common factors since we would have canceled them out beforehand. Two integers without common factors can not both be even, so at least one of p, q is odd.

Case 1. p is odd. Then p^2 is odd while $2q^2$ is not. Therefore $p^2 \neq 2q^2$.

Case 2. p is even and q is odd. Then p^2 is divisible by 4 while $2q^2$ is not. Therefore $p^2 \neq 2q^2$. □

The set $\mathbb{Q}$ of rational numbers is incomplete. It has "gaps," one of which occurs at $\sqrt{2}$. These gaps are really more like pinholes; they have zero width. Incompleteness is what is *wrong* with $\mathbb{Q}$. Our goal is to complete $\mathbb{Q}$ by filling in its gaps. An elegant method to arrive at this goal is **Dedekind cuts** in which one visualizes real numbers as places at which a line may be cut with scissors. See Figure 3.

Definition A **cut** in $\mathbb{Q}$ is a pair of subsets A, B of $\mathbb{Q}$ such that

(a) $A \cup B = \mathbb{Q}$, $A \neq \emptyset$, $B \neq \emptyset$, $A \cap B = \emptyset$.

(b) If $a \in A$ and $b \in B$ then $a < b$.

(c) A contains no largest element.

[†] A subtler fact that you may find useful is the prime factorization theorem mentioned above. Any integer ≥ 2 can be factored into a product of prime numbers. For example, 120 is the product of primes $2 \cdot 2 \cdot 2 \cdot 3 \cdot 5$. Prime factorization is unique except for the order in which the factors appear. An easy consequence is that if a prime number p divides an integer k and if k is the product mn of integers then p divides m or it divides n. After all, by uniqueness, the prime factorization of k is just the product of the prime factorizations of m and n.

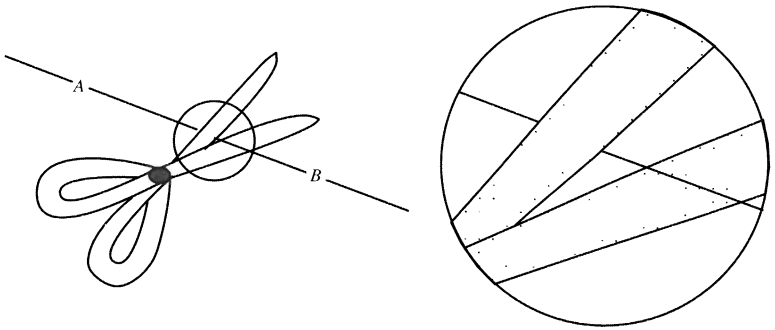


Figure 3 A Dedekind cut.

A is the left-hand part of the cut and B is the right-hand part. We denote the cut as $x = A|B$. Making a semantic leap, we now answer the question "what is a real number?"

Definition A **real number** is a cut in $\mathbb{Q}$.

$\mathbb{R}$ is the class[†] of all real numbers $x = A|B$. We will show that in a natural way $\mathbb{R}$ is a complete ordered field containing $\mathbb{Q}$. Before spelling out what this means, here are two examples of cuts.

(i) $A|B = \{r \in \mathbb{Q} : r < 1\} \mid \{r \in \mathbb{Q} : r \geq 1\}$.

(ii) $A|B = \{r \in \mathbb{Q} : r \leq 0 \text{ or } r^2 < 2\} \mid \{r \in \mathbb{Q} : r > 0 \text{ and } r^2 \geq 2\}$.

It is convenient to say that $A|B$ is a **rational cut** if it is like the cut in (i): for some fixed rational number c, A is the set of all rationals $< c$ while B is the rest of $\mathbb{Q}$. The B-set of a rational cut contains a smallest element c, and conversely, if $A|B$ is a cut in $\mathbb{Q}$ and B contains a smallest element c then $A|B$ is the rational cut at c. We write c^* for the rational cut at c. This lets us think of $\mathbb{Q} \subset \mathbb{R}$ by identifying c with c^*. It is like thinking of $\mathbb{Z}$ as a subset of $\mathbb{Q}$ since the integer n in $\mathbb{Z}$ can be thought of as the fraction $n/1$ in $\mathbb{Q}$. In the same way the rational number c in $\mathbb{Q}$ can be thought of as the cut at c. It is just a different way of looking at c. It is in this sense that we write

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

There is an order relation $x \leq y$ on cuts that fairly cries out for attention.

[†] The word "class" is used instead of the word "set" to emphasize that for now the members of $\mathbb{R}$ are set-pairs $A|B$, and not the numbers that belong to A or B. The notation $A|B$ could be shortened to A since B is just the rest of $\mathbb{Q}$. We write $A|B$, however, as a mnemonic device. It *looks* like a cut.

Definition The cut $x = A|B$ is **less than or equal** to the cut $y = C|D$ if $A \subset C$.

We write $x \leq y$ if x is less than or equal to y and we write $x < y$ if $x \leq y$ and $x \neq y$. If $x = A|B$ is less than $y = C|D$ then $A \subset C$ and $A \neq C$, so there is some $c_0 \in C \setminus A$. Since the A-set of a cut contains no largest element, there is also a $c_1 \in C$ with $c_0 < c_1$. All the rational numbers c with $c_0 \leq c \leq c_1$ belong to $C \setminus A$. Thus, $x < y$ implies that not only is $C \setminus A$ non-empty, but it contains infinitely many elements.

The property distinguishing $\mathbb{R}$ from $\mathbb{Q}$ and which is at the bottom of every significant theorem about $\mathbb{R}$ involves upper bounds and least upper bounds; or equivalently, lower bounds and greatest lower bounds.

$M \in \mathbb{R}$ is an **upper bound** for a set $S \subset \mathbb{R}$ if each $s \in S$ satisfies

$$s \leq M.$$

We also say that the set S is **bounded above** by M. An upper bound for S that is less than all other upper bounds for S is a **least upper bound** for S. The least upper bound for S is denoted l.u.b. (S). For example,

3 is an upper bound for the set of negative integers.

-1 is the least upper bound for the set of negative integers.

1 is the least upper bound for the set

$$\{x \in \mathbb{Q} : \exists n \in \mathbb{N} \text{ and } x = 1 - 1/n\}.$$

-100 is an upper bound for the empty set.

A least upper bound for S may or may not belong to S. This is why you should say "least upper bound *for* S" rather than "least upper bound *of* S."

2 Theorem *The set $\mathbb{R}$, constructed by means of Dedekind cuts, is* ***complete***[†] *in the sense that it satisfies the* ***Least Upper Bound Property****:*

If S is a non-empty subset of $\mathbb{R}$ and is bounded above then in $\mathbb{R}$ there exists a least upper bound for S.

Proof Easy! Let $\mathscr{C} \subset \mathbb{R}$ be any non-empty collection of cuts which is bounded above, say by the cut $X|Y$. Define

$$C = \{a \in \mathbb{Q} : \text{ for some cut } A|B \in \mathscr{C}, a \in A\} \text{ and } D = \text{ the rest of } \mathbb{Q}.$$

It is easy to see that $z = C|D$ is a cut. Clearly, it is an upper bound for $\mathscr{C}$ since the "A" for every element of $\mathscr{C}$ is contained in C. Let $z' = C'|D'$

[†] There is another, related, sense in which $\mathbb{R}$ is complete. See Theorem 5 below.

be any upper bound for $\mathscr{C}$. By the assumption that $A|B \leq C'|D'$ for all $A|B \in \mathscr{C}$, we see that the "A" for every member of $\mathscr{C}$ is contained in C'. Hence $C \subset C'$, so $z \leq z'$. That is, among all upper bounds for $\mathscr{C}$, z is least. □

The simplicity of this proof is what makes cuts good. We go from $\mathbb{Q}$ to $\mathbb{R}$ by pure thought. To be more complete, as it were, we describe the natural arithmetic of cuts. Let cuts $x = A|B$ and $y = C|D$ be given. How do we add them? subtract them? ... Generally the answer is to do the corresponding operation to the elements comprising the two halves of the cuts, being careful about negative numbers. The sum of x and y is $x + y = E|F$ where

$$\begin{aligned} E &= \{r \in \mathbb{Q} : \text{ for some } a \in A \text{ and } c \in C, r = a + c\} \\ F &= \text{ the rest of } \mathbb{Q}. \end{aligned}$$

It is easy to see that $E|F$ is a cut in $\mathbb{Q}$ and that it doesn't depend on the order in which x and y appear. That is, cut addition is well defined and $x + y = y + x$. The zero cut is 0^* and $0^* + x = x$ for all $x \in \mathbb{R}$. The additive inverse of $x = A|B$ is $-x = C|D$ where

$$\begin{aligned} C &= \{r \in \mathbb{Q} : \text{ for some } b \in B, \text{ not the smallest element of } B, r = -b\} \\ D &= \text{ the rest of } \mathbb{Q}. \end{aligned}$$

Then $(-x) + x = 0^*$. Correspondingly, the difference of cuts is $x - y = x + (-y)$. Another property of cut addition is **associativity**:

$$(x + y) + z = x + (y + z).$$

This follows from the corresponding property of $\mathbb{Q}$.

Multiplication is trickier to define. It helps to first say that the cut $x = A|B$ is positive if $0^* < x$ or negative if $x < 0^*$. Since 0 lies in A or B, a cut is either positive, negative, or zero. If $x = A|B$ and $y = C|D$ are nonnegative cuts then their product is $x \cdot y = E|F$ where

$$\begin{aligned} E = \{r \in \mathbb{Q} \ : \ & r \leq 0 \text{ or } \exists a \in A, \\ & \text{such that } a > 0, c > 0, \text{ and } r = ac\}, \end{aligned}$$

and F is the rest of $\mathbb{Q}$. If x is positive and y is negative then we define the product to be $-(x \cdot (-y))$. Since x and $-y$ are both positive cuts this makes sense and is a negative cut. Similarly, if x is negative and y is positive then by definition their product is the negative cut $-((-x) \cdot y)$, while if x and y are both negative then their product is the positive cut $(-x) \cdot (-y)$.

Verifying the arithmetic properties for multiplication is tedious, to say the least, and somehow nothing seems to be gained by writing out every detail. (To pursue cut arithmetic further you could read Landau's classically boring book, *Foundations of Analysis*.) To get the flavor of it, let's the check the commutativity of multiplication: $x \cdot y = y \cdot x$ for cuts $x = A|B$, $y = C|D$. If x, y are positive then

$$\{ac : a \in A, c \in C, a > 0, c > 0\} = \{ca : c \in C, a \in A, c > 0, a > 0\}$$

implies that $x \cdot y = y \cdot x$. If x is positive and y is negative then

$$x \cdot y = -(x \cdot (-y)) = -((-y) \cdot x) = y \cdot x.$$

The second equality holds because we have already checked commutativity for positive cuts. The remaining two cases are checked similarly. There are eight cases to check for associativity and eight more for distributivity. All are simple and we omit their proofs. The real point is that cut arithmetic can be defined and it satisfies the same field properties that $\mathbb{Q}$ does:

The operation of cut addition is
well defined, natural, commutative, associative, and
has inverses with respect to the neutral element 0^*.
The operation of cut multiplication
is well defined, natural, commutative, associative,
distributive over cut addition, and has inverses of
nonzero elements with respect to the neutral element 1^*.

By definition, a **field** is a system consisting of a set of elements and two operations, addition and multiplication, that have the preceding algebraic properties – commutativity, associativity, etc. Besides just existing, cut arithmetic is consistent with $\mathbb{Q}$ arithmetic in the sense that if $c, r \in \mathbb{Q}$ then $c^* + r^* = (c + r)^*$ and $c^* \cdot r^* = (cr)^*$. By definition, this is what we mean when we say that $\mathbb{Q}$ is a **subfield** of $\mathbb{R}$. The cut order enjoys the additional properties of

transitivity. $x < y < z$ implies $x < z$.

trichotomy. Either $x < y$, $y < x$, or $x = y$, but only one of the three things is true.

translation. $x < y$ implies $x + z < y + z$.

By definition, this is what we mean when we say that $\mathbb{R}$ is an **ordered field**. Besides, the product of positive cuts is positive and cut order is consistent with $\mathbb{Q}$ order: $c^* < r^*$ if and only if $c < r$ in $\mathbb{Q}$. By definition, this is what we mean when we say that $\mathbb{Q}$ is an ordered subfield of $\mathbb{R}$. To summarize

3 Theorem *The set $\mathbb{R}$ of all cuts in $\mathbb{Q}$ is a complete ordered field that contains $\mathbb{Q}$ as an ordered subfield.*

The **magnitude** or absolute value of $x \in \mathbb{R}$ is

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0. \end{cases}$$

Thus, $x \leq |x|$. A basic, constantly used fact about magnitude is the following.

4 Triangle Inequality *For all $x, y \in \mathbb{R}$, $|x + y| \leq |x| + |y|$.*

Proof The translation and transitivity properties of the order relation imply that adding y and $-y$ to the inequalities $x \leq |x|$ and $-x \leq |x|$ gives

$$\begin{aligned} x + y &\leq |x| + y \leq |x| + |y| \\ -x - y &\leq |x| - y \leq |x| + |y|. \end{aligned}$$

Since $x + y \leq |x| + |y|$ and $-(x + y) \leq |x| + |y|$, we infer that $|x + y| \leq |x| + |y|$ as asserted. □

Next, suppose we try the same cut construction in $\mathbb{R}$ that we did in $\mathbb{Q}$. Are there gaps in $\mathbb{R}$ that can be detected by cutting $\mathbb{R}$ with scissors? The natural definition of a cut in $\mathbb{R}$ is a division $\mathcal{A}|\mathcal{B}$ where $\mathcal{A}$ and $\mathcal{B}$ are disjoint, non-empty subcollections of $\mathbb{R}$ with $\mathcal{A} \cup \mathcal{B} = \mathbb{R}$, and $a < b$ for all $a \in \mathcal{A}$, $b \in \mathcal{B}$. Further, $\mathcal{A}$ contains no largest element. Now, each $b \in \mathcal{B}$ is an upper bound for $\mathcal{A}$. Therefore $y = \text{l. u. b.}(\mathcal{A})$ exists and $a \leq y \leq b$ for all $a \in \mathcal{A}$ and $b \in \mathcal{B}$. By trichotomy,

$$\mathcal{A}|\mathcal{B} = \{x \in \mathbb{R} : x < y\} \mid \{x \in \mathbb{R} : x \geq y\}.$$

In other words, $\mathbb{R}$ has no gaps. Every cut in $\mathbb{R}$ occurs exactly *at* a real number.

Allied to the existence of $\mathbb{R}$ is its uniqueness. Any complete ordered field $\mathbb{F}$ containing $\mathbb{Q}$ as an ordered subfield corresponds to $\mathbb{R}$ in a way preserving all the ordered field structure. To see this, take any $\varphi \in \mathbb{F}$ and associate to it the cut $A|B$ where

$$A = \{r \in \mathbb{Q} : r < \varphi \text{ in } \mathbb{F}\}.$$

This correspondence makes $\mathbb{F}$ equivalent to $\mathbb{R}$.

Upshot The real number system $\mathbb{R}$ exists and it satisfies the properties of a complete ordered field; the properties are not assumed as axioms, but are proved by logically analyzing the Dedekind construction of $\mathbb{R}$. Having gone through all this cut rigmarole, it must be remarked that it is a rare working mathematician who actually thinks of $\mathbb{R}$ as a complete ordered field or as the set of all cuts in $\mathbb{Q}$. Rather, he or she thinks of $\mathbb{R}$ as points on the x-axis, just as in calculus. You too should picture $\mathbb{R}$ this way, the only benefit of the cut derivation being that you should now unhesitatingly accept the least upper bound property of $\mathbb{R}$ as a true fact.

Note $\pm\infty$ are not real numbers since $\mathbb{Q}|\emptyset$ and $\emptyset|\mathbb{Q}$ are not cuts. Although some mathematicians think of $\mathbb{R}$ together with $-\infty$ and $+\infty$ as an "extended real number system," it is simpler to leave well enough alone and just deal with $\mathbb{R}$ itself. Nevertheless, it is convenient to write expressions like "$x \to \infty$" to indicate that a real variable x grows larger and larger without bound.

If S is a non-empty subset of $\mathbb{R}$ then its **supremum** is its least upper bound when S is bounded above and is said to be $+\infty$ otherwise; its **infimum** is its greatest lower bound when S is bounded below and is said to be $-\infty$ otherwise. (In Exercise 17 you are asked to invent the notion of greatest lower bound.) By definition the supremum of the empty set is $-\infty$. This is reasonable, considering that every real number, no matter how negative, is an upper bound for $\emptyset$, and the least upper bound should be as far leftward as possible, namely $-\infty$. Similarly, the infimum of the empty set is $+\infty$. We write $\sup S$ and $\inf S$ for the supremum and infimum of S.

Cauchy sequences

As mentioned above there is a second sense in which $\mathbb{R}$ is complete. It involves the concept of convergent sequences. Let $a_1, a_2, a_3, a_4, \cdots = (a_n)$, $n \in \mathbb{N}$, be a sequence of real numbers. The sequence (a_n) **converges to the limit** $b \in \mathbb{R}$ as $n \to \infty$ provided that for each $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$,

$$|a_n - b| < \epsilon.$$

The statistician's language is evocative here. Think of $n = 1, 2, \ldots$ as a sequence of times and say that the sequence (a_n) converges to b provided that *eventually* all its terms nearly equal b. In symbols,

$$\forall \epsilon > 0 \ \exists N \in \mathbb{N} \text{ such that } n \geq N \Rightarrow |a_n - b| < \epsilon.$$

If the limit b exists it is not hard to see that it is unique, and we write

$$\lim_{n\to\infty} a_n = b \text{ or } a_n \to b.$$

Suppose that $\lim_{n\to\infty} a_n = b$. Since all the numbers a_n are eventually near b they are all near each other; i.e., every convergent sequence obeys a **Cauchy condition**:

$$\forall \epsilon > 0 \ \exists N \in \mathbb{N} \text{ such that } n, m \geq N \Rightarrow |a_n - a_m| < \epsilon.$$

The converse of this fact is a fundamental property of $\mathbb{R}$.

5 Theorem $\mathbb{R}$ *is* ***complete*** *with respect to Cauchy sequences in the sense that if* (a_n) *is a sequence of real numbers obeying a Cauchy condition then it converges to a limit in* $\mathbb{R}$.

Proof Let A be the set of real numbers comprising the sequence (a_n),

$$A = \{x \in \mathbb{R} : \exists n \in \mathbb{N} \text{ and } a_n = x\}.$$

We first observe that A is a bounded set in $\mathbb{R}$. Taking $\epsilon = 1$ in the Cauchy condition, there is an integer N_1 such that for all $n, m \geq N_1$, $|a_n - a_m| < 1$. Then, for each $n \geq N_1$

$$|a_n - a_{N_1}| < 1. \tag{10}$$

Clearly the finite set $a_1, a_2, \ldots, a_{N_1}, a_{N_1} - 1, a_{N_1} + 1$ is bounded, (any finite set is bounded); say all its elements belong to the interval $[-M, M]$. According to (10), $[-M, M]$ contains A so A is bounded.

Next, consider the set

$$S = \{s \in [-M, M] : \exists \text{ infinitely many } n \in \mathbb{N}, \text{ for which } a_n \geq s\}.$$

That is, $a_n \geq s$ infinitely often. Clearly $-M \in S$ and S is bounded above by M. According to the least upper bound property of $\mathbb{R}$ there exists $b \in \mathbb{R}$, $b = \text{l. u. b. } S$. We claim that the sequence (a_n) converges to b.

Given $\epsilon > 0$ we must show that there exists an N such that for all $n \geq N$, $|a_n - b| < \epsilon$. The Cauchy condition provides an N_2 such that

$$m, n \geq N_2 \quad \Rightarrow \quad |a_m - a_n| < \frac{\epsilon}{2}. \tag{11}$$

All elements of S are $\leq b$, so the larger number $b + \epsilon/2$ does not belong to S. Only finitely often does a_n exceed $b + \epsilon/2$. That is, for some $N_3 \geq N_2$,

$$n \geq N_3 \quad \Rightarrow \quad a_n \leq b + \frac{\epsilon}{2}.$$

Since b is a least upper bound for S, the smaller number $b - \epsilon/2$ can not also be an upper bound for S. Some $s \in S$ is $\geq b - \epsilon/2$, which implies that $a_n \geq s > b - \epsilon/2$ infinitely often. In particular, there exists $N \geq N_3$ such that $a_N > b - \epsilon/2$. Since $N \geq N_3$, $a_N < b + \epsilon/2$ and

$$a_N \in (b - \epsilon/2, \ b + \epsilon/2).$$

Since $N \geq N_2$, (11) implies

$$|a_n - b| \leq |a_n - a_N| + |a_N - b| < \epsilon,$$

which verifies convergence. □

Restating Theorem 5 gives the

6 Cauchy Convergence Criterion for sequences *A sequence* (a_n) *in* $\mathbb{R}$ *converges if and only if*

$$\forall \epsilon > 0 \ \exists N \in \mathbb{N} \textit{ such that } n, m \geq N \Rightarrow |a_n - a_m| < \epsilon.$$

Further description of $\mathbb{R}$

The elements of $\mathbb{R} \setminus \mathbb{Q}$ are **irrational numbers**. If x is irrational and r is rational then $y = x + r$ is irrational. For if y is rational then so is $y - r = x$, the difference of rationals being rational. Similarly, if $r \neq 0$ then rx is irrational. It follows that the reciprocal of an irrational number is irrational. From these observations we will show that the rational and irrational numbers are thoroughly mixed up with each other.

Let $a < b$ be given in $\mathbb{R}$. Define the **intervals** (a, b) and $[a, b]$ as

$$(a, b) = \{x \in \mathbb{R} : a < x < b\}$$
$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}.$$

7 Theorem *Every interval* (a, b)*, no matter how small, contains both rational and irrational numbers.*

Proof This is certainly true of the interval $(0, 1)$ since it contains the numbers $1/2$ and $1/\sqrt{2}$. For the general interval (a, b), think of a, b as cuts $a = A|A'$, $b = B|B'$. The fact that $a < b$ implies the set $B \setminus A$ contains two distinct rational numbers, say r, s. Thus $a \leq r < s \leq b$. The transformation

$$T : t \mapsto r + (s - r)t$$

sends the interval $(0, 1)$ to the interval (r, s). Since r, s, and $s-r$ are rational, T sends rationals to rationals and irrationals to irrationals. That is, (r, s) contains both rationals and irrationals, and so does the larger interval (a, b). □

Theorem 7 expresses the fact that between any two rational numbers lies an irrational number; and between any two irrational numbers lies a rational number. This is a fact worth thinking about for it seems implausible at first. Spend some time trying to picture the situation, especially in light of the following related facts:

(a) There is no first (i.e., smallest) rational number in the interval $(0, 1)$.
(b) There is no first irrational number in the interval $(0, 1)$.
(c) There are strictly more irrational numbers in the interval $(0, 1)$ (in the cardinality sense explained in Section 4) than there are rational numbers.

The transformation in the proof of Theorem 7 shows that the real line is like rubber: stretch it out and it never breaks.

A somewhat obscure and trivial fact about $\mathbb{R}$ is its **Archimedean property**: for each $x \in \mathbb{R}$ there is an integer n that is greater than x. In other words, there exist arbitrarily large integers. The Archimedean property is true for $\mathbb{Q}$ since $p/q \leq |p|$. It follows that it is true for $\mathbb{R}$. Given $x = A|B$, just choose a rational number $r \in B$ and an integer $n > r$. Then $n > x$. An equivalent way to state the Archimedean property is that there exist arbitrarily small reciprocals of integers.

Mildly interesting is the existence of ordered fields for which the Archimedean property fails. One example is the field $\mathbb{R}(x)$ of rational functions with real coefficients. Each such function is of the form

$$R(x) = \frac{p(x)}{q(x)}$$

where p and q are polynomials with real coefficients and q is not the zero polynomial. (It does not matter that $q(x) = 0$ at a finite number of points.) Addition and multiplication are defined in the usual fashion of high school algebra, and it is easy to see that $\mathbb{R}(x)$ is a field. The order relation on $\mathbb{R}(x)$ is also easy to define. If $R(x) > 0$ for all sufficiently large x then we say that R is positive in $\mathbb{R}(x)$, and if $R - S$ is positive then we write $S < R$. Since a nonzero rational function vanishes (has value zero) at only finitely many $x \in \mathbb{R}$, we get trichotomy: either $R = S$, $R < S$, or $S < R$. (To be rigorous, we need to prove that the values of a rational function do not change sign for x large enough.) The other order properties are equally easy to check, and $\mathbb{R}(x)$ is an ordered field.

Is $\mathbb{R}(x)$ Archimedean? That is, given $R \in \mathbb{R}(x)$, does there exist a natural number $n \in \mathbb{R}(x)$ such that $R < n$? (A number n is the rational function whose numerator is the constant polynomial $p(x) = n$, a polynomial of degree zero, and whose denominator is the constant polynomial $q(x) = 1$.) The answer is "no." Take $R(x) = x/1$. The numerator is x and the denominator is 1. Clearly, we have $n < x$, not the opposite, so $\mathbb{R}(x)$ fails to be Archimedean.

The same remarks hold for any positive rational function $R = p(x)/q(x)$ where the degree of p exceeds the degree of q. In $\mathbb{R}(x)$, R is never less than a natural number. (You might ask yourself: exactly which rational functions are less than n?)

The ϵ-principle

Finally let us note a nearly trivial principle that turns out to be invaluable in deriving inequalities and equalities in $\mathbb{R}$.

8 Theorem (ϵ-principle) *If a, b are real numbers and if for each $\epsilon > 0$, $a \leq b + \epsilon$, then $a \leq b$. If x, y are real numbers and for each $\epsilon > 0$, $|x - y| \leq \epsilon$, then $x = y$.*

Proof Trichotomy implies that either $a \leq b$ or $a > b$. In the latter case we can choose ϵ, $0 < \epsilon < a - b$ and get the absurdity

$$\epsilon < a - b \leq \epsilon.$$

Hence $a \leq b$. Similarly, if $x \neq y$ then choosing ϵ, $0 < \epsilon < |x - y|$ gives the contradiction $\epsilon < |x - y| \leq \epsilon$. Hence $x = y$. See also Exercise 11. □

3 Euclidean Space

Given sets A and B, the **Cartesian product** of A and B is the set $A \times B$ of all ordered pairs (a, b) such that $a \in A$ and $b \in B$. (The name comes from Descartes who pioneered the idea of the (x, y)-coordinate system in geometry.) See Figure 4.

The Cartesian product of $\mathbb{R}$ with itself m times is denoted $\mathbb{R}^m$. Elements of $\mathbb{R}^m$ are vectors, ordered m-tuples of real numbers, $(x_1, \ldots, x_m)$. In this terminology, real numbers are called scalars and $\mathbb{R}$ is called the scalar field. When vectors are added, subtracted, and multiplied by scalars according to the rules

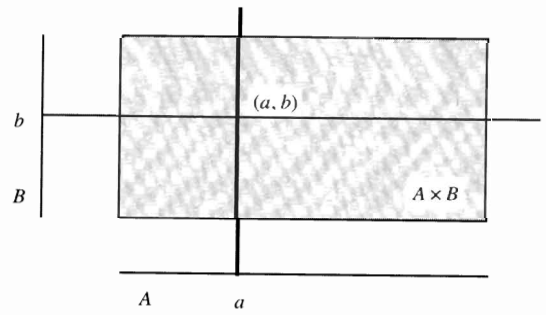


Figure 4 The Cartesian product $A \times B$.

$$(x_1, \ldots, x_m) + (y_1, \ldots, y_m) = (x_1 + y_1, \ldots, x_m + y_m)$$
$$(x_1, \ldots, x_m) - (y_1, \ldots, y_m) = (x_1 - y_1, \ldots, x_m - y_m)$$
$$c\,(x_1, \ldots, x_m) = (cx_1, \ldots, cx_m)$$

then these operations obey the natural laws of linear algebra: commutativity, associativity, etc. There is another operation defined on $\mathbb{R}^m$, the **dot product** (also called the scalar product or inner product). The dot product of $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_m)$ is

$$\langle x, y \rangle = x_1 y_1 + \cdots + x_m y_m.$$

Remember: the dot product of two vectors is a scalar, not a vector. The dot product operation is bilinear, symmetric, and positive definite; i.e., for any $x, y, z \in \mathbb{R}^m$ and any $c \in \mathbb{R}$,

$$\langle x, y + cz \rangle = \langle x, y \rangle + c\langle x, z \rangle,$$
$$\langle x, y \rangle = \langle y, x \rangle,$$
$$\langle x, x \rangle \geq 0 \text{ and } \langle x, x \rangle = 0 \text{ if and only if } x \text{ is the zero vector.}$$

The **length** or **magnitude** of a vector $x \in \mathbb{R}^m$ is defined to be

$$|x| = \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + \cdots + x_m^2}.$$

See Exercise 15 which legalizes taking roots. Expressed in coordinate-free language, the basic fact about the dot product is the

9 Cauchy-Schwarz Inequality *For all* $x, y \in \mathbb{R}^m$, $\langle x, y \rangle \leq |x|\,|y|$.

Proof Tricky! For any vectors x, y consider the new vector $w = x + ty$ where $t \in \mathbb{R}$ is a varying scalar. Then

$$f(t) = \langle w, w \rangle = \langle x + ty, x + ty \rangle$$

is a real-valued function of t. In fact, $f(t) \geq 0$ since the dot product of any vector with itself is nonnegative. The bilinearity properties of the dot product imply that

$$f(t) = \langle x, x \rangle + 2t\langle x, y \rangle + t^2\langle y, y \rangle = c + bt + at^2$$

is a quadratic function of t. Nonnegative quadratic functions of $t \in \mathbb{R}$ have nonpositive discriminants, $b^2 - 4ac \leq 0$. For if $b^2 - 4ac > 0$ then $f(t)$ has two real roots, between which $f(t)$ is negative. See Figure 5.

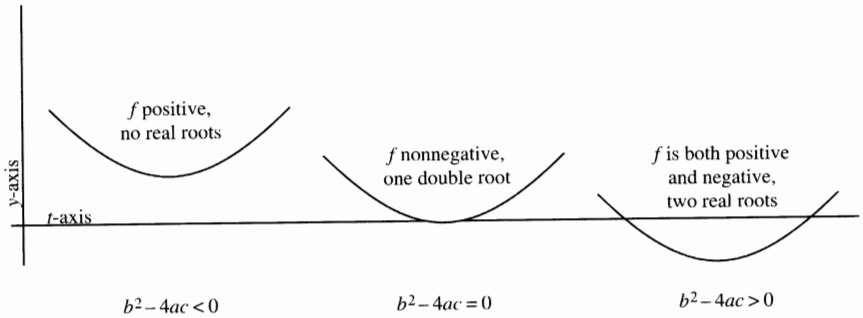


Figure 5 Quadratic graphs.

But $b^2 - 4ac \leq 0$ means that

$$4\langle x, y \rangle^2 - 4\langle x, x \rangle\langle y, y \rangle \leq 0.$$

Therefore $\langle x, y \rangle \leq \sqrt{\langle x, x \rangle}\sqrt{\langle y, y \rangle} = |x|\,|y|$. □

The Cauchy-Schwarz inequality implies easily the **Triangle Inequality for vectors:** For all $x, y \in \mathbb{R}^m$,

$$|x + y| \leq |x| + |y|.$$

For $|x + y|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + 2\langle x, y \rangle + \langle y, y \rangle$. By Cauchy-Schwarz, $2\langle x, y \rangle \leq 2\,|x|\,|y|$. Thus,

$$|x + y|^2 \leq |x|^2 + 2\,|x|\,|y| + |y|^2 = (|x| + |y|)^2.$$

Taking the square root of both sides gives the result.

The **Euclidean distance** between vectors $x, y \in \mathbb{R}^m$ is defined as the length of their difference,

$$|x - y| = \sqrt{\langle x - y,\ x - y\rangle} = \sqrt{(x_1 - y_1)^2 + \cdots + (x_m - y_m)^2}.$$

From the Triangle Inequality for vectors follows the **Triangle Inequality for distance.** For all $x, y, z \in \mathbb{R}^m$,

$$|x - z| \le |x - y| + |y - z|.$$

To prove it, think of $x - z$ as a vector sum $(x - y) + (y - z)$ and apply the Triangle Inequality for vectors. See Figure 6.

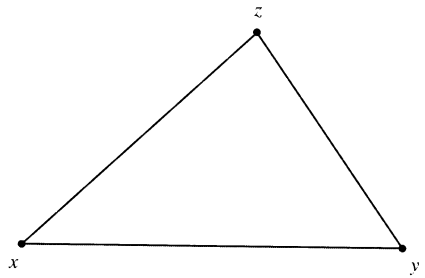


Figure 6 How the Triangle Inequality gets its name.

Geometric intuition in Euclidean space can carry you a long way in real analysis, especially in being able to forecast whether a given statement is true or not. Your geometric intuition will grow with experience and contemplation. We begin with some vocabulary.

The j^{th} coordinate of the point $(x_1, \ldots, x_m)$ is the number x_j appearing in the j^{th} position. The j^{th} coordinate axis is the set of points $x \in \mathbb{R}^m$ whose k^{th} coordinates are zero for all $k \neq j$. The origin of $\mathbb{R}^m$ is the zero vector, $(0, \ldots, 0)$. The **first orthant** of $\mathbb{R}^m$ is the set of points $x \in \mathbb{R}^m$ all of whose coordinates are nonnegative. When $m = 2$, the first orthant is the first quadrant. The **integer lattice** is the set $\mathbb{Z}^m \subset \mathbb{R}^m$ of ordered m-tuples of integers. The integer lattice is also called the integer grid. See Figure 7.

A **box** is a Cartesian product of intervals

$$[a_1, b_1] \times \cdots \times [a_m, b_m]$$

in $\mathbb{R}^m$. (A box is also called a rectangular parallelpiped.) The **unit cube** in $\mathbb{R}^m$ is the box $[0, 1]^m = [0, 1] \times \cdots \times [0, 1]$. See Figure 8.

The **unit ball** and **unit sphere** in $\mathbb{R}^m$ are the sets

$$B^m = \{x \in \mathbb{R}^m : |x| \le 1\}$$
$$S^{m-1} = \{x \in \mathbb{R}^m : |x| = 1\}.$$

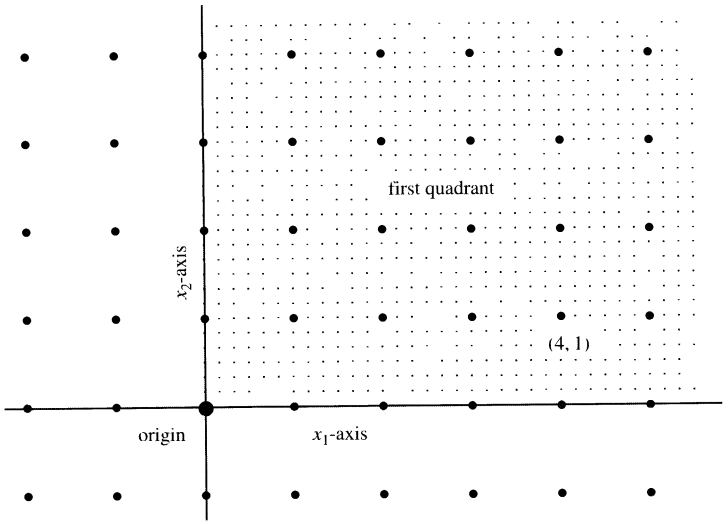


Figure 7 The integer lattice and first quadrant.

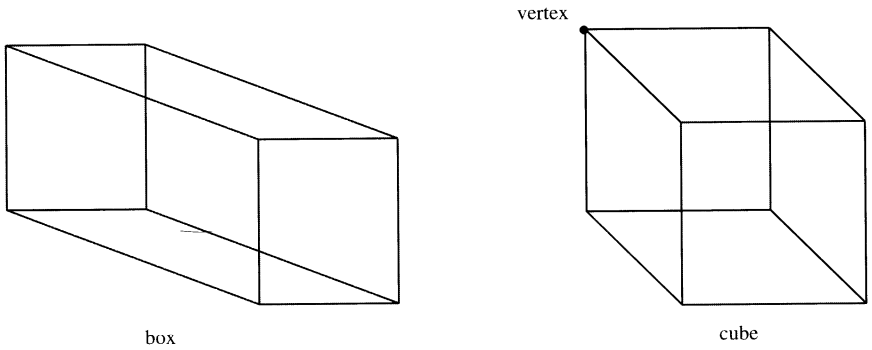


Figure 8 A box and a cube.

The reason for the exponent $m - 1$ is that the sphere is $(m - 1)$-dimensional as an object in its own right, although it does *live* in m-space. In 3-space, the surface of a ball is a two-dimensional film, the 2-sphere S^2. See Figure 9.

A set $E \subset \mathbb{R}^m$ is **convex** if for each pair of points $x, y \in E$, the straight line segment between x and y is also contained in E. The unit ball is an example of a convex set. To see this, take any two points in B^m and draw the segment between them. It obviously lies in B^m. See Figure 10.

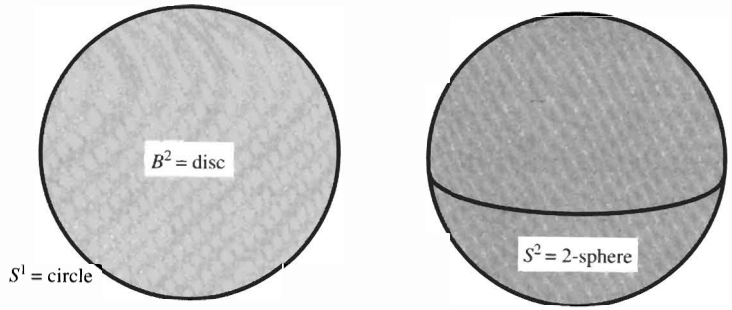


Figure 9 A 2-disc with its boundary circle, and a 2-sphere.

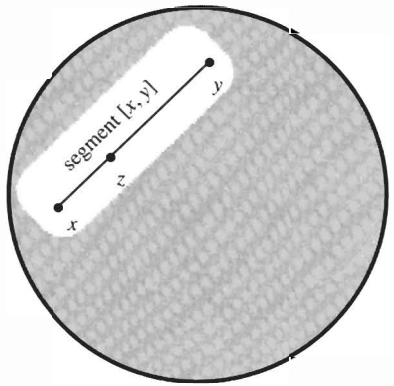


Figure 10 Convexity of the ball.

To give a mathematical proof, it is useful to describe the line segment between x and y with a formula. The straight line determined by distinct points $x, y \in \mathbb{R}^m$ is the set of all linear combinations $sx+ty$ where $s+t = 1$, and the line segment is the set of linear combinations where s and t are ≤ 1. Such linear combinations $sx + ty$ with $s + t = 1$ and $0 \leq s, t \leq 1$ are called **convex combinations**. The line segment is denoted as $[x, y]$. (This notation is consistent with the interval notation $[a, b]$. See Exercise 25.) Now if $x, y \in B^m$ and $sx + ty = z$ is a convex combination of x and y then, using the Cauchy-Schwarz inequality and the fact that $2st \geq 0$, we get

$$\begin{aligned}\langle z, z\rangle &= s^2\langle x, x\rangle + 2st\langle x, y\rangle + t^2\langle y, y\rangle \\ &\leq s^2|x|^2 + 2st|x||y| + t^2|y|^2 \\ &\leq s^2 + 2st + t^2 = (s+t)^2 = 1.\end{aligned}$$

Taking the square root of both sides gives $|z| \leq 1$, which proves convexity of the ball.

Inner product spaces

An **inner product** on a vector space V is an operation $\langle\ ,\ \rangle$ on pairs of vectors in V that satisfies the same conditions that the dot product in Euclidean space does, namely, bilinearity, symmetry, and positive definiteness. A vector space equipped with an inner product is an **inner product space**. The discriminant proof of the Cauchy-Schwarz Inequality is valid for any inner product defined on any real vector space, even if the space is infinite-dimensional and the standard coordinate proof would make no sense. For the discriminant proof uses only the inner product properties, and not the particular definition of the dot product in Euclidean space.

$\mathbb{R}^m$ has dimension m because it has a basis $e_1, \ldots, e_m$. Other vector spaces are more general. For example, let $C([a, b], \mathbb{R})$ denote the set of all of continuous real valued functions defined on the interval $[a, b]$. (See Section 6 or your old calculus book for the definition of continuity.) It is a vector space in a natural way, the sum of continuous functions being continuous and the scalar multiple of a continuous function being continuous. The vector space, however, has no finite basis. It is infinite-dimensional. Even so, there is a natural inner product,

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx.$$

Cauchy-Schwarz applies to this inner product, just as to any inner product, and we infer a general integral inequality valid for any two continuous functions,

$$\int_a^b f(x)g(x)\,dx \quad \leq \quad \sqrt{\int_a^b f(x)^2 dx}\,\sqrt{\int_a^b g(x)^2\,dx}.$$

It would be challenging to prove such an inequality from scratch, would it not?

A **norm** on a vector space V is any function $\|\ \| : V \to \mathbb{R}$ with the three properties of vector length: namely, if $v, w \in V$ and $\lambda \in \mathbb{R}$ then

$\|v\| \geq 0$ and $\|v\| = 0$ if and only if $v = 0$,
$\|\lambda v\| = |\lambda| \, \|v\|$,
$\|v + w\| \leq \|v\| + \|w\|$.

An inner product $\langle \ , \ \rangle$ defines a norm as $\|v\| = \sqrt{\langle v, v \rangle}$, but not all norms come from inner products. The unit sphere $\{v \in V : \langle v, v \rangle = 1\}$ for every inner product is smooth (has no corners) while for the norm

$$\|v\|_{\max} = \max\{|v_1|, |v_2|\}$$

defined on $v = (v_1, v_2) \in \mathbb{R}^2$, the unit sphere is the perimeter of the square $\{(v_1, v_2) \in \mathbb{R}^2 : |v_1| \leq 1 \text{ and } |v_2| \leq 1\}$. It has corners and so it does not arise from an inner product. See Exercises 43, 44, and the Manhattan metric on page 72.

The simplest Euclidean space beyond $\mathbb{R}$ is the plane $\mathbb{R}^2$. Its xy-coordinates can be used to define a multiplication,

$$(x, y) \bullet (x', y') = (xx' - yy', xy' + x'y).$$

The point $(1, 0)$ corresponds to the multiplicative unit element 1, while the point $(0, 1)$ corresponds to $i = \sqrt{-1}$, which converts the plane to the field $\mathbb{C}$ of complex numbers. Complex analysis is the study of functions of a complex variable, i.e., functions $f(z)$ where z and $f(z)$ lie in $\mathbb{C}$. Complex analysis is the good twin and real analysis the evil one: beautiful formulas and elegant theorems seem to blossom spontaneously in the complex domain, while toil and pathology rule the reals. Nevertheless, complex analysis relies more on real analysis than the other way around.

4 Cardinality

Let A and B be sets. A **function** $f : A \to B$ is a rule or mechanism which, when presented with any element $a \in A$, produces an element $b = f(a)$ of B. It need not be defined by a formula. Think of a function as a device into which you feed elements of A and out of which pour elements of B. See Figure 11.

We also call f a mapping or a map or a transformation. The set A is the **domain** of the function and B is the **target**. The **range** or **image** of f is the subset of the target,

$$\{b \in B : \text{ there exists at least one element } a \in A \text{ with } f(a) = b\}.$$

See Figure 12.

Try to write f instead of $f(x)$ to denote a function. The function is the device which when confronted with input x produces output $f(x)$. The function is the device, not the output.

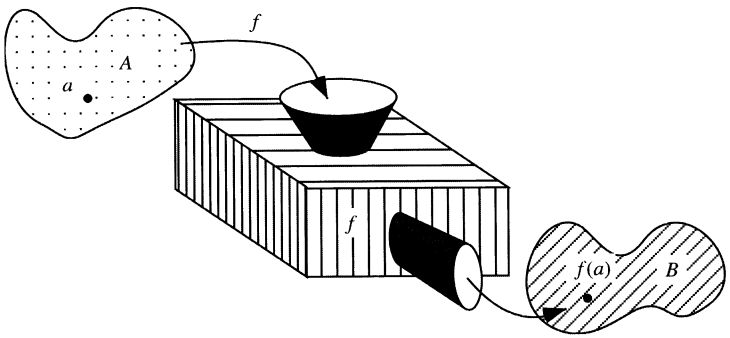


Figure 11 The function f as a machine.

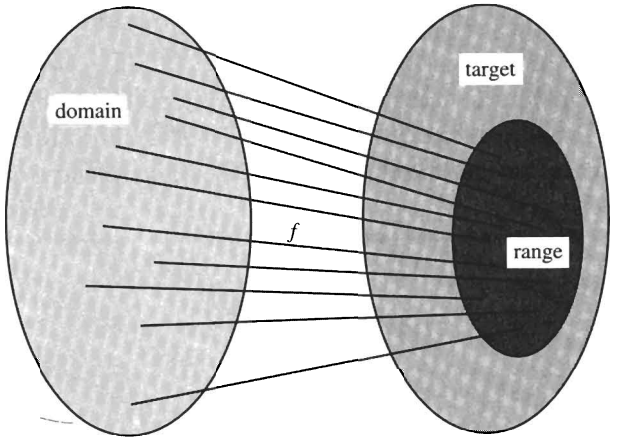


Figure 12 The domain, target, and range of a function.

Think also of a function dynamically. At time zero all the elements of A are sitting peacefully in A. Then the function applies itself to them and throws them into B. At time one all the elements that were formerly in A are now transferred into B. Each $a \in A$ gets sent to some element $f(a) \in B$.

A mapping $f : A \to B$ is an **injection** (or is **one-to-one**) if for each pair of distinct elements $a, a' \in A$, the elements $f(a), f(a')$ are distinct in B. That is,

$$a \neq a' \quad \Rightarrow \quad f(a) \neq f(a').$$

The mapping f is a **surjection** (or is **onto**) if for each $b \in B$ there is at least one $a \in A$ such that $f(a) = b$. That is, the range of f is B.

A mapping is a **bijection** if it is both injective and surjective. It is one-to-one and onto. If $f : A \to B$ is a bijection then the inverse map $f^{-1} : B \to A$ is a bijection where $f^{-1}(b)$ is by definition the unique element $a \in A$ such that $f(a) = b$.

The **identity map** of any set to itself is the bijection that takes each $a \in A$ and sends it to itself, $\mathrm{id}(a) = a$.

If $f : A \to B$ and $g : B \to C$ then the **composite** $g \circ f : A \to C$ is the function that sends $a \in A$ to $g(f(a)) \in C$. If f and g are injective then so is $g \circ f$, while if f and g are surjective then so is $g \circ f$.

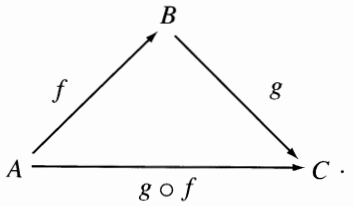
In particular the composite of bijections is a bijection. If there is a bijection from A onto B then A and B are said to have **equal cardinality**,[†] and we write $A \sim B$. The relation $\sim$ is an equivalence relation. That is,

(a) $A \sim A$.

(b) $A \sim B$ implies $B \sim A$.

(c) $A \sim B \sim C$ implies $A \sim C$.

(a) follows from the fact that the identity map bijects A to itself. (b) follows from the fact that the inverse of a bijection $A \to B$ is a bijection $B \to A$. (c) follows from the fact that the composite of bijections f and g is a bijection $g \circ f$,



A set S is

finite if it is empty or for some $n \in \mathbb{N}$, $S \sim \{1, \ldots, n\}$.

infinite if it is not finite.

denumerable if $S \sim \mathbb{N}$.

countable if it is finite or denumerable.

uncountable if it is not countable.

We also write $\mathrm{card}\, A = \mathrm{card}\, B$ and $\#A = \#B$ when A, B have equal cardinality.

[†] The word "cardinal" indicates the number of elements in the set. The cardinal numbers are $0, 1, 2, \ldots$ The first infinite cardinal number is **aleph null**, $\aleph_0$. One says the $\mathbb{N}$ has $\aleph_0$ elements. A mystery of math is the **Continuum Hypothesis** which states that $\mathbb{R}$ has cardinality $\aleph_1$, the second infinite cardinal. Equivalently, if $\mathbb{N} \subset S \subset \mathbb{R}$, the Continuum Hypothesis asserts that $S \sim \mathbb{N}$ or $S \sim \mathbb{R}$. No intermediate cardinalities exist. You can pursue this issue in Paul Cohen's book, *Set Theory and the Continuum Hypothesis.*

If S is denumerable then there is a bijection $f : \mathbb{N} \to S$, and this gives a way to list the elements of S as $s_1 = f(1), s_2 = f(2), s_3 = f(3)$, etc. Conversely, if a set S is presented as an infinite list (without repetition) $S = \{s_1, s_2, s_3, \ldots\}$ then it is denumerable: define $f(k) = s_k$ for $k \in \mathbb{N}$. In brief, denumerable = listable.

Let's begin with a truly remarkable cardinality result, that although $\mathbb{N}$ and $\mathbb{R}$ are both infinite, $\mathbb{R}$ is more infinite than $\mathbb{N}$. Namely,

10 Theorem $\mathbb{R}$ *is uncountable.*

Proof There are other proofs of the uncountability of $\mathbb{R}$, but none so beautiful as this one. It is due to Cantor. I assume that you accept the fact that each real number x has a decimal expansion, $x = N.x_1x_2x_3\ldots$, and it is uniquely determined by x if one agrees never to terminate the expansion with an infinite string of 9's. (See also Exercise 16.) We want to prove that $\mathbb{R}$ is uncountable. Suppose it is not uncountable. It is countable and, being infinite, it must be denumerable. Accordingly let $f : \mathbb{N} \to \mathbb{R}$ be a bijection. Using f, we list the elements of $\mathbb{R}$ along with their decimal expansions as an array, and consider the digits x_{ii} that occur along the diagonal in this array. See Figure 13.

$$
\begin{array}{lcl}
f(1) & = & N_1\ x_{11}\ x_{12}\ x_{13}\ x_{14}\ x_{15}\ x_{16}\ x_{17} \\
f(2) & = & N_2\ x_{21}\ x_{22}\ x_{23}\ x_{24}\ x_{25}\ x_{26}\ x_{27} \\
f(3) & = & N_3\ x_{31}\ x_{32}\ x_{33}\ x_{34}\ x_{35}\ x_{36}\ x_{37} \\
f(4) & = & N_4\ x_{41}\ x_{42}\ x_{43}\ x_{44}\ x_{45}\ x_{46}\ x_{47} \\
f(5) & = & N_5\ x_{51}\ x_{52}\ x_{53}\ x_{54}\ x_{55}\ x_{56}\ x_{57} \\
f(6) & = & N_6\ x_{61}\ x_{62}\ x_{63}\ x_{64}\ x_{65}\ x_{66}\ x_{67} \\
f(7) & = & N_7\ x_{71}\ x_{72}\ x_{73}\ x_{74}\ x_{75}\ x_{76}\ x_{77}
\end{array}
$$

Figure 13 Cantor's diagonal method.

For each i, choose a digit y_i such that $y_i \neq x_{ii}$ and $y_i \neq 9$. Where is the number $y = 0.y_1y_2y_3\ldots$? Is it $f(1)$? No, because the first digit in the decimal expansion of $f(1)$ is x_{11} and $y_1 \neq x_{11}$. Is it $f(2)$? No, because the second digit in the decimal expansion of $f(2)$ is x_{22} and $y_2 \neq x_{22}$. Is it $f(k)$? No, because the k^{th} digit in the decimal expansion of $f(k)$ is x_{kk} and $y_k \neq x_{kk}$. Nowhere in the list do we find y. Nowhere! Therefore the list could not account for every real number, and $\mathbb{R}$ must have been uncountable. □

11 Corollary *$[a, b]$ and (a, b) are uncountable.*

Proof There are bijections from (a, b) onto $(-1, 1)$ onto the unit semicircle onto $\mathbb{R}$ shown in Figure 14.

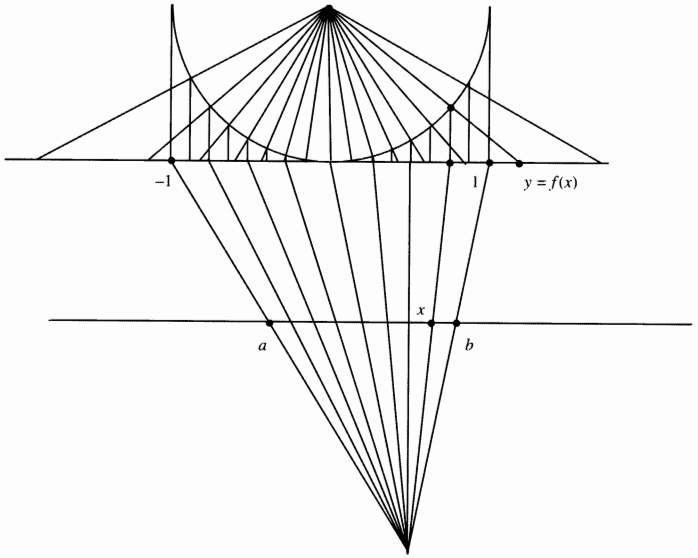


Figure 14 Equicardinality of (a, b), $(-1, 1)$, and $\mathbb{R}$.

The composite f bijects (a, b) onto $\mathbb{R}$, so (a, b) is uncountable. Since $[a, b]$ contains (a, b), it too is uncountable. □

The remaining results in this section are of a more positive flavor.

12 Theorem *Each infinite set S contains a denumerable subset.*

Proof Since S is infinite it is non-empty and contains an element s_1. Since S is infinite the set $S \setminus \{s_1\} = \{s \in S : s \neq s_1\}$ is non-empty and there exists $s_2 \in S \setminus \{s_1\}$. Since S is an infinite set, $S \setminus \{s_1, s_2\} = \{s \in S : s \neq s_1, s_2\}$ is non-empty and there exists $s_3 \in S \setminus \{s_1, s_2\}$. Continuing this way gives a list (s_n) of distinct elements of S. The set of these elements forms a denumerable subset of S. □

13 Theorem *An infinite subset A of a denumerable set B is denumerable.*

Proof There exists a bijection $f : \mathbb{N} \to B$. Each element of A appears exactly once in the list $f(1), f(2), f(3), \ldots$ of B. Define $g(k)$ to be the k^{th} element of A appearing in the list. Since A is infinite, $g(k)$ is defined for all $k \in \mathbb{N}$. Thus $g : \mathbb{N} \to A$ is a bijection and A is denumerable. □

14 Corollary *The sets of even integers and of prime integers are denumerable.*

Proof They are infinite subsets of $\mathbb{N}$, which is denumerable. □

15 Theorem *$\mathbb{N} \times \mathbb{N}$ is denumerable.*

Proof Think of $\mathbb{N} \times \mathbb{N}$ as an $\infty \times \infty$ matrix and walk along the successive counter-diagonals. See Figure 15. This gives a list

$(1, 1), (2, 1), (1, 2), (3, 1), (2, 2), (1, 3), (4, 1), (3, 2), (2, 3), (1, 4), (5, 1), \ldots$

of $\mathbb{N} \times \mathbb{N}$ and proves that $\mathbb{N} \times \mathbb{N}$ is denumerable. □

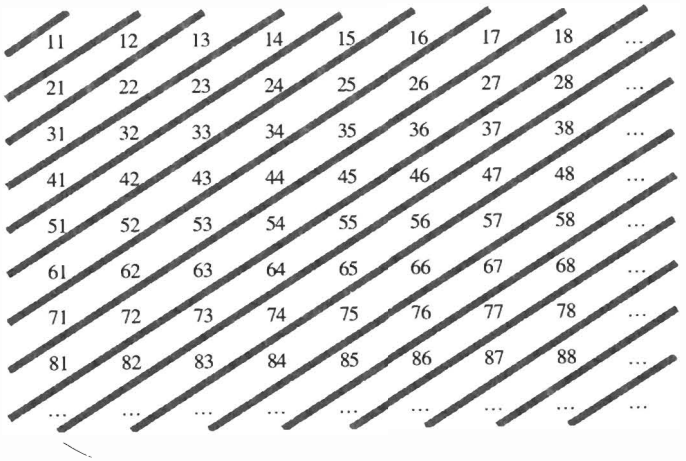


Figure 15 Counter-diagonals in an $\infty \times \infty$ matrix.

16 Corollary *The Cartesian product of denumerable sets A and B is denumerable.*

Proof $\mathbb{N} \sim \mathbb{N} \times \mathbb{N} \sim A \times B$. □

17 Theorem *If $f : \mathbb{N} \to B$ is a surjection and B is infinite then B is denumerable.*

Proof For each $b \in B$, the set $\{k \in \mathbb{N} : f(k) = b\}$ is non-empty and hence contains a smallest element, say $h(b) = k$ is the smallest integer that is sent to b by f. Clearly, if $b \neq b'$ then $h(b) \neq h(b')$. That is, $h : B \to \mathbb{N}$ is an injection which bijects B to $hB \subset \mathbb{N}$. Since B is infinite, so is hB. By Theorem 13, hB is denumerable and therefore so is B. □

18 Corollary *The denumerable union of denumerable sets is denumerable.*

Proof Suppose that $A_1, A_2, \ldots$ is a sequence of denumerable sets. List the elements of A_i as $a_{i1}, a_{i2}, \ldots$ and define

$$f : \mathbb{N} \times \mathbb{N} \to A = \bigcup A_i$$
$$(i, j) \mapsto a_{ij}$$

Clearly f is a surjection. According to Theorem 15, there is a bijection $g : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$. The composite $f \circ g$ is a surjection $\mathbb{N} \to A$. Since A is infinite, Theorem 17 implies it is denumerable. □

19 Corollary *$\mathbb{Q}$ is denumerable.*

Proof $\mathbb{Q}$ is the denumerable union of the denumerable sets $A_q = \{p/q : p \in \mathbb{Z}\}$ as q ranges over $\mathbb{N}$. □

20 Corollary *For each $m \in \mathbb{N}$, $\mathbb{Q}^m$ is denumerable.*

Proof Apply the induction principle. If $m = 1$, then the previous corollary states that $\mathbb{Q}^1$ is denumerable. Knowing inductively that $\mathbb{Q}^{m-1}$ is denumerable and $\mathbb{Q}^m = \mathbb{Q}^{m-1} \times \mathbb{Q}$, the result follows from Corollary 16. □

Combination laws for countable sets are similar to those for denumerable sets. As is easily checked,

Every subset of a countable set is countable.
A countable set that contains a denumerable subset is denumerable.
The Cartesian product of finitely many countable sets is countable.
The countable union of countable sets is countable.

5* Comparing Cardinalities

The following result gives a way to conclude that two sets have the same cardinality. Roughly speaking the condition is that card $A \leq$ card B and card $B \leq$ card A.

21 Schroeder-Bernstein Theorem *If A, B are sets and $f : A \to B$, $g : B \to A$ are injections then there exists a bijection $h : A \to B$.*

Proof-sketch Consider the dynamic Venn diagram, Figure 16. The disc labeled gfA is the image of A under the map $g \circ f$. It is a subset of A. The ring between A and gfA divides into two subrings. A_0 is the set of points in A that do not lie in the image of g, while A_1 is the set points in the image of g that do not lie in gfA. Similarly, B_0 is the set of points in B that do not lie in fA, while B_1 is the set of points in fA that do not lie in fgB. There is a natural bijection h from the pair of rings $A_0 \cup A_1 = A \setminus gfA$ to the pair of rings $B_0 \cup B_1 = B \setminus fgB$. It equals f on the outer ring $A_0 = A \setminus gB$

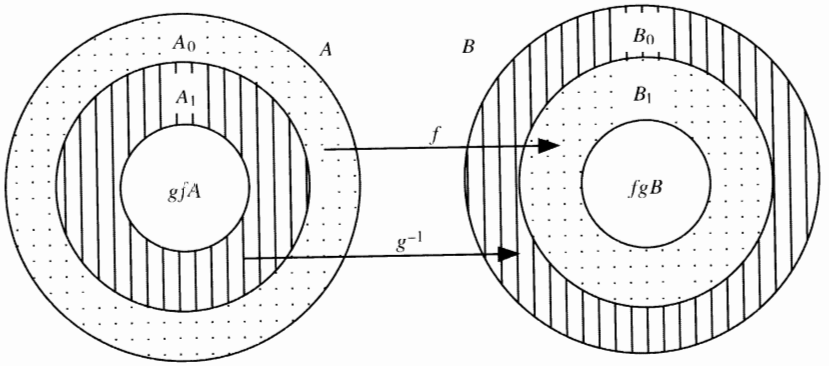


Figure 16 Pictorial proof of the Schroeder-Bernstein Theorem.

and it is g^{-1} on the inner ring $A_1 = gB \setminus gfA$. (The map g^{-1} is not defined on all of A, but it is defined on the set gB.) In this notation, h sends A_0 onto B_1 and sends A_1 onto B_0. It switches the indices. Repeat this on the next pair of rings for A and B. That is, look at gfA instead of A and fgB instead of B. The next two rings in A, B are

$$A_2 = gfA \setminus gfgB \qquad A_3 = gfgB \setminus gfgfA$$
$$B_2 = fgB \setminus fgfA \qquad B_3 = fgfA \setminus fgfgB.$$

Send A_2 to B_3 by f and A_3 to B_2 by g^{-1}. The rings A_i are disjoint, and so are the rings B_i, so repetition gives a bijection

$$\phi : \bigsqcup A_i \to \bigsqcup B_i,$$

($\bigsqcup$ indicates disjoint union) defined by

$$\phi(x) = \begin{cases} f(x) & \text{if } x \in A_i \text{ and } i \text{ is even} \\ g^{-1}(x) & \text{if } x \in A_i \text{ and } i \text{ is odd} \end{cases}$$

Let $A_* = A \setminus (\bigcup A_i)$ and $B_* = B \setminus (\bigcup B_i)$ be the rest of A and B. Then f bijects A_* to B_*, and ϕ extends to a bijection $h : A \to B$ defined by

$$h(x) = \begin{cases} \phi(x) & \text{if } x \in \bigcup A_i \\ f(x) & \text{if } x \in A_*. \end{cases}$$ □

A supplementary aid in understanding the Schroeder Bernstein proof is the following crossed ladder diagram, Figure 17.

Exercise 33 asks you to show directly that $(a, b) \sim [a, b]$. This makes sense since $(a, b) \subset [a, b] \subset \mathbb{R}$ and $(a, b) \sim \mathbb{R}$ should certainly imply that $(a, b) \sim [a, b] \sim \mathbb{R}$.

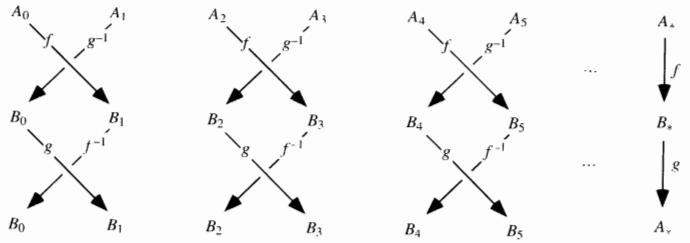


Figure 17 Diagramatic proof of the Schroeder-Bernstein Theorem.

The Schroeder-Bernstein theorem gives a quick, indirect solution to the exercise. The inclusion map $i : (a, b) \hookrightarrow [a, b]$ sending x to x injects (a, b) into $[a, b]$, while the function $j(x) = x/2 + (a + b)/4$ injects $[a, b]$ into (a, b). The existence of the two injections implies by the Schroeder-Bernstein Theorem that there is a bijection $(a, b) \sim [a, b]$.

6* The Skeleton of Calculus

The behavior of a continuous function defined on an interval $[a, b]$ is at the root of all calculus theory. Using solely the l.u.b. property of the real numbers we rigorously derive the basic properties of such functions. The function $f : [a, b] \to \mathbb{R}$ is **continuous** if for each $\epsilon > 0$ and each $x \in [a, b]$ there is a $\delta > 0$ such that

$$t \in [a, b] \text{ and } |t - x| < \delta \quad \Rightarrow \quad |f(t) - f(x)| < \epsilon.$$

See Figure 18.

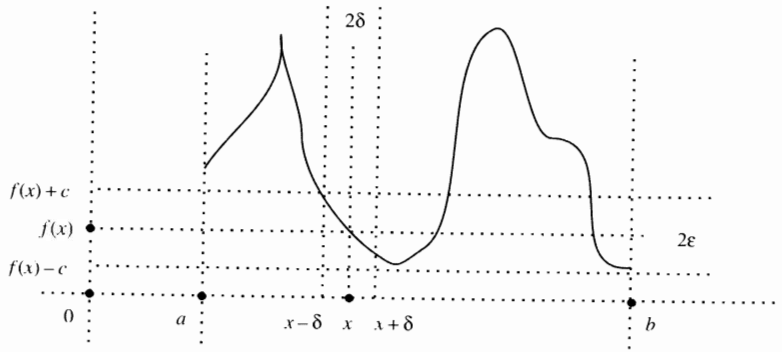


Figure 18 The graph of a continuous function of a real variable.

Continuous functions are found everywhere in analysis and topology. Theorems 22, 23, 24 present their simplest properties. Later we generalize these results to functions that are neither real valued nor dependent on a real variable. Although it is possible to give a combined proof of Theorems 22 and 23, I prefer to highlight the l.u.b. property and keep them separate.

22 Theorem *The values of a continuous function defined on an interval* $[a, b]$ *form a bounded subset of* $\mathbb{R}$. *That is, there exist* $m, M \in \mathbb{R}$ *such that for all* $x \in [a, b]$, $m \leq f(x) \leq M$.

Proof For $x \in [a, b]$, let V_x be the value set of $f(t)$ as t varies from a to x,

$$V_x = \{y \in \mathbb{R} : \text{ for some } t \in [a, x], \ y = f(t)\}.$$

Set

$$X = \{x \in [a, b] : V_x \text{ is a bounded subset of } \mathbb{R}\}.$$

We must prove that $b \in X$. Clearly $a \in X$ and b is an upper bound for X. Since X is non-empty and bounded above, there exists in $\mathbb{R}$ a least upper bound $c \leq b$ for X. Take $\epsilon = 1$ in the definition of continuity. There exists a $\delta > 0$ such that $|x - c| < \delta$ implies $|f(x) - f(c)| < 1$. Since c is the least upper bound for X, there exists $x \in X$ in the interval $[c - \delta, c]$. (Otherwise $c - \delta$ is a smaller upper bound for X.) Now as t varies from a to c, the value $f(t)$ varies first in the bounded set V_x and then in the bounded set $J = (f(c) - 1, f(c) + 1)$. See Figure 19.

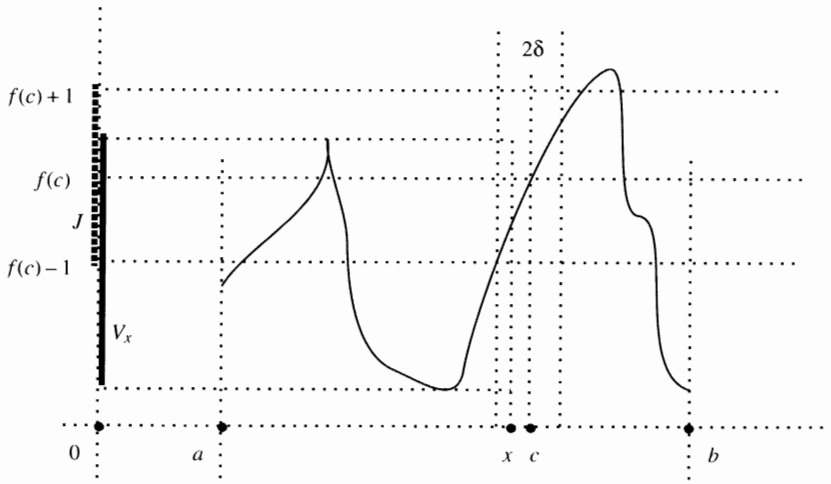


Figure 19 The value set V_x and the interval J.

The union of two bounded sets is a bounded set and it follows that V_c is bounded, so $c \in X$. Besides, if $c < b$ then $f(t)$ continues to vary in the bounded set J for $t > c$, contrary to the fact that c is an upper bound for X. Thus, $c = b$, $b \in X$, and the values of f form a bounded subset of $\mathbb{R}$. □

23 Theorem *A continuous function f defined on an interval $[a, b]$ takes on absolute minimum and absolute maximum values: for some $x_0, x_1 \in [a, b]$ and for all $x \in [a, b]$,*

$$f(x_0) \leq f(x) \leq f(x_1).$$

Proof Let $M = \text{l. u. b. } f(t)$ as t varies in $[a, b]$. By Theorem 22, M exists. Consider the set $X = \{x \in [a, b] : \text{l. u. b. } V_x < M\}$ where, as above, V_x is the set of values of $f(t)$ as t varies on $[a, x]$.

Case 1. $f(a) = M$. Then f takes on a maximum at a and the theorem is proved.

Case 2. $f(a) < M$. Then $X \neq \emptyset$ and we can consider the least upper bound of X, say c. If $f(c) < M$, we choose $\epsilon > 0$ with $\epsilon < M - f(c)$. By continuity, there exists a $\delta > 0$ such that $|t - c| < \delta$ implies $|f(t) - f(c)| < \epsilon$. Thus, l. u. b. $V_c < M$. If $c < b$ this implies that there exist points t to the right of c at which l. u. b. $V_t < M$, contrary to the fact that c is an upper bound of such points. Therefore, $c = b$ which implies that $M < M$, a contradiction. Having arrived at a contradiction from the supposition that $f(c) < M$, we duly conclude that $f(c) = M$, so f assumes a maximum at c. The situation with minima is similar. □

24 Intermediate Value Theorem *A continuous function defined on an interval $[a, b]$ takes on (or "achieves," "assumes," or "attains") all intermediate values: if $f(a) = \alpha$, $f(b) = \beta$, and γ is given, $\alpha \leq \gamma \leq \beta$, then there is some $c \in [a, b]$ such that $f(c) = \gamma$. The same conclusion holds if $\beta \leq \gamma \leq \alpha$.*

The theorem is pictorially obvious. A continuous function has a graph that is a curve without break points. Such a graph can not jump from one height to another. It must pass through all intermediate heights.

Proof Set $X = \{x \in [a, b] : \text{l. u. b. } V_x \leq \gamma\}$ and $c = \text{l. u. b. } X$. Now c exists because X is non-empty (it contains a) and it is bounded above (by b). We claim that $f(c) = \gamma$, as shown in Figure 20.

To prove it, we just eliminate the other two possibilities, $f(c) < \gamma$ and $f(c) > \gamma$, by showing that each leads to a contradiction. Suppose that $f(c) < \gamma$ and take $\epsilon = \gamma - f(c)$. Continuity gives $\delta > 0$ such that $|t - c| < \delta$ implies $|f(t) - f(c)| < \epsilon$. That is,

$$t \in (c - \delta, c + \delta) \quad \Rightarrow \quad f(t) < \gamma,$$

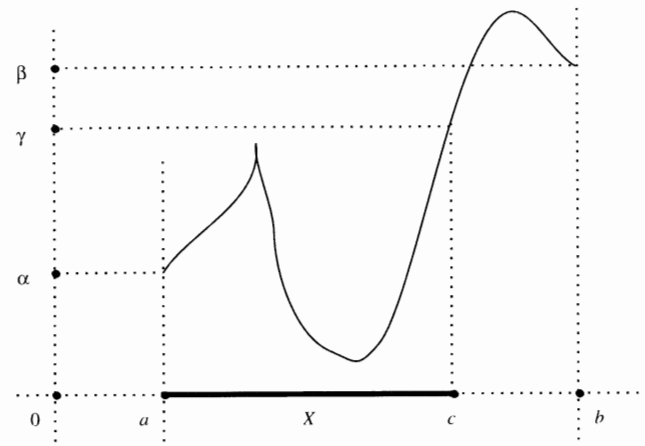


Figure 20 $x \in X$ implies that $f(x) \leq \gamma$.

so $c + \delta/2 \in X$, contrary to c being an upper bound of X.

Suppose that $f(c) > \gamma$ and take $\epsilon = f(c) - \gamma$. Continuity gives $\delta > 0$ such that $|t - c| < \delta$ implies $|f(t) - f(c)| < \epsilon$. That is,

$$t \in (c - \delta, c + \delta) \Rightarrow f(t) > \gamma,$$

so $c - \delta/2$ is an upper bound for X, contrary to c being the least upper bound for X.

Since $f(c)$ is neither $< \gamma$ nor $> \gamma$, we get $f(c) = \gamma$. □

A combination of Theorems 22, 23, 24 and Exercise 40 could well be called the

Fundamental Theorem of Continuous Functions *Every continuous real valued function of a real variable $x \in [a, b]$ is bounded, achieves minimum, intermediate, and maximum values, and is uniformly continuous.*

Exercises

I have adopted Moe Hirsch's star system for the exercises. One star is hard, two stars is very hard, and a three-star exercise is a question to which I do not know the answer.

0. Prove that for all sets A, B, C the formula

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

is true. Here is the solution written out in gory detail. *Imitate this style in writing out proofs in this course.*
Hypothesis. A, B, C are sets.
Conclusion. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
Proof. To prove two sets are equal we must show that every element of the first set is an element of the second set and vice versa. Referring to Figure 21, let x denote an element of the set $A \cap (B \cup C)$. It belongs

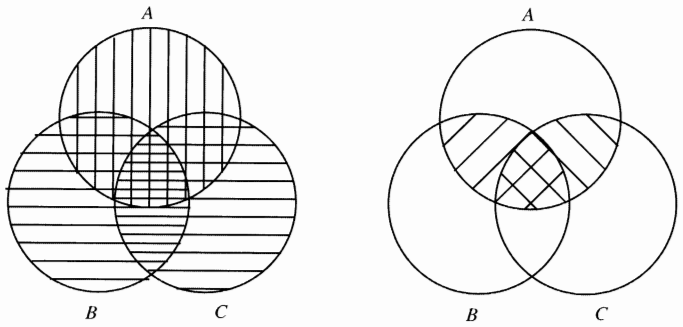


Figure 21 A is ruled vertically, B and C are ruled horizontally, $A \cap B$ is ruled diagonally, and $A \cap C$ is ruled counter-diagonally.

to A and it belongs to B or to C. Therefore x belongs to $A \cap B$ or it belongs to $A \cap C$. Thus x belongs to the set $(A \cap B) \cup (A \cap C)$ and we have shown that every element of the first set $A \cap (B \cup C)$ belongs to the second set $(A \cap B) \cup (A \cap C)$.
On the other hand let y denote an element of the set $(A \cap B) \cup (A \cap C)$. It belongs to $A \cap B$ or it belongs to $A \cap C$. Therefore it belongs to A and it belongs to B or to C. Thus y belongs to $A \cap (B \cup C)$ and we have shown that every element of the second set $(A \cap B) \cup (A \cap C)$ belongs to the first set $A \cap (B \cup C)$.
Since each element of the first set belongs to the second set and each element of the second belongs to the first, the two sets are equal, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. QED

1. Prove that for all sets A, B, C the formula

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

is true.

2. If several sets $A, B, C, \ldots$ all are subsets of the same set X then the differences $X \setminus A$, $X \setminus B$, $X \setminus C$, ...are the **complements** of $A, B, C, \ldots$ in X and are denoted $A^c, B^c, C^c, \ldots$. The symbol A^c is read "A complement."
   (a) Prove that $(A^c)^c = A$.
   (b) Prove **deMorgan's Law**: $(A \cap B)^c = A^c \cup B^c$ and derive from it the law $(A \cup B)^c = A^c \cap B^c$.
   (c) Draw Venn diagrams to illustrate the two laws.
   (d) Generalize these laws to more than two sets.
3. Recast the following English sentences in mathematics, using correct mathematical grammar. Preserve their meaning.
   (a) 2 is the smallest prime number.
   (b) The area of any bounded plane region is bisected by some line parallel to the x-axis.
   *(c) "All that glitters is not gold."
*4. What makes the following sentence ambiguous? "A death row prisoner can't have too much hope."
5. Negate the following sentences in English using correct mathematical grammar.
   (a) If roses are red, violets are blue.
   *(b) He will sink unless he swims.
6. Why is the square of an odd integer odd and the square of an even integer even? What is the situation for higher powers? [Hint: Prime factorization.]
7. (a) Why does 4 divide every even integer square?
   (b) Why does 8 divide every even integer cube?
   (c) Why can 8 never divide twice an odd cube?
   (d) Prove that the cube root of 2 is irrational.
8. Suppose that the integer k is not a perfect n^{th} power.
   (a) Prove that $\sqrt[n]{k} \notin \mathbb{Q}$.
   (b) Infer that the n^{th} root of an integer is either an integer or it is irrational. It is never a fraction.
9. Let $x = A|B, x' = A'|B'$ be cuts in $\mathbb{Q}$. We defined

$$x + x' = (A + A') \mid \text{rest of } \mathbb{Q}.$$

(a) Show that although $B + B'$ is disjoint from $A + A'$, it may happen in degenerate cases that $\mathbb{Q}$ is not the union of $A + A'$ and $B + B'$.
(b) Infer that the definition of $x + x'$ as $(A + A')|(B + B')$ would be incorrect.
(c) Why did we not define $x \cdot x' = (A \cdot A')|$rest of $\mathbb{Q}$?

10. A multiplicative inverse of a nonzero cut $x = A|B$ is a cut $y = C|D$ such that $x \cdot y = 1^*$.
(a) If $x > 0^*$, what are C and D?
(b) If $x < 0^*$, what are they?
(c) Prove that x uniquely determines y.

11. Prove that there exists no smallest positive real number. Does there exist a smallest positive rational number? Given a real number x, does there exist a smallest real number $y > x$?

12. Let $b =$ l. u. b. S, where S is a bounded nonempty subset of $\mathbb{R}$.
(a) Given $\epsilon > 0$ show that there exists an $s \in S$ with

$$b - \epsilon \leq s \leq b.$$

(b) Can $s \in S$ always be found so that $b - \epsilon < s < b$?
(c) If $x = A|B$ is a cut in $\mathbb{Q}$, show that $x =$ l. u. b. A.

13. Prove that $\sqrt{2} \in \mathbb{R}$ by showing that $x \cdot x = 2$ where $x = A|B$ is the cut in $\mathbb{Q}$ with $A = \{r = \mathbb{Q} : r \leq 0 \text{ or } r^2 < 2\}$. [Hint: Use Exercise 12. See also Exercise 15, below.]

14. Given $y \in \mathbb{R}$, $n \in \mathbb{N}$, and $\epsilon > 0$, show that for some $\delta > 0$, if $|u - y| < \delta$ then $|u^n - y^n| < \epsilon$. [Hint: Prove the inequality when $n = 1, n = 2$, and then do induction on n using the identity

$$u^n - y^n = (u - y)(u^{n-1} + u^{n-2}y + \cdots + y^{n-1}).]$$

15. Given $x > 0$ and $n \in \mathbb{N}$, prove that there is a unique $y > 0$ such that $y^n = x$. That is, $y = \sqrt[n]{x}$ exists and is unique. [Hint: Consider

$$y = \text{l. u. b.}\{s \in \mathbb{R} : s^n \leq x\}.$$

Then use Exercise 14 to show that y^n can be neither $< x$ nor $> x$.]

16. Prove that real numbers correspond bijectively to decimal expansions not terminating in an infinite strings of 9's, as follows. The decimal expansion of $x \in \mathbb{R}$ is $N.x_1x_2\ldots$ where N is the largest integer $\leq x$, x_1 is the largest integer $\leq 10(x - N)$, x_2 is the largest integer $\leq 100(x - (N + x_1/10))$, and so on.

(a) Show that each x_k is a digit between 0 and 9.
(b) Show that for each k there is an $\ell \geq k$ such that $x_\ell \neq 9$.
(c) Conversely, show that for each such expansion $N.x_1x_2\ldots$ not terminating in an infinite string of 9's, the set

$$\{N,\ N + \frac{x_1}{10},\ N + \frac{x_1}{10} + \frac{x_2}{100},\ \ldots\}$$

is bounded and its least upper bound is a real number x with decimal expansion $N.x_1x_2\ldots$.
(d) Repeat the exercise with a general base in place of 10.

17. Formulate the definition of the greatest lower bound (g.l.b.) of a set of real numbers. State a g.l.b. property of $\mathbb{R}$ and show it is equivalent to the l.u.b. property of $\mathbb{R}$.

18. Let $f : A \to B$ be a function. That is, f is some rule or device which, when presented with any element $a \in A$, produces an element $b = f(a)$ of B. The **graph** of f is the set S of all pairs $(a, b) \in A \times B$ such that $b = f(a)$.
(a) If you are given a subset $S \subset A \times B$, how can you tell if it is the graph of some function? (That is, what are the set theoretic properties of a graph?)
(b) Let $g : B \to C$ be a second function and consider the composed function $g \circ f : A \to C$. Assume that $A = B = C = [0, 1]$, draw $A \times B \times C$ as the unit cube in 3-space, and try to relate the graphs of f, g, and $g \circ f$ in the cube.

19. A **fixed point** of a function $f : A \to A$ is a point $a \in A$ such that $f(a) = a$. The **diagonal** of $A \times A$ is the set of all pairs (a, a) in $A \times A$.
(a) Show that $f : A \to A$ has a fixed point if and only if the graph of f intersects the diagonal.
(b) Prove that every continuous function $f : [0, 1] \to [0, 1]$ has at least one fixed point.
(c) Is the same true for continuous functions $f : (0, 1) \to (0, 1)$?†
(d) Is the same true for discontinuous functions?

20. Given a cube in $\mathbb{R}^m$, what is the largest ball it contains? Given a ball in $\mathbb{R}^m$, what is the largest cube it contains? What are the largest ball and cube contained in a given box in $\mathbb{R}^m$?

† A question posed in this manner means that, as well as answering the question with a "yes" or a "no," you should give a proof if your answer is "yes" or a specific counter-example if your answer is "no." Also, to do this exercise you should read Theorems 22, 23, 24.

21. A rational number p/q is **dyadic** if q is a power of 2, $q = 2^k$ for some nonnegative integer k. For example 0, 3/8, 3/1, $-3/256$, are dyadic rationals, but $1/3, 5/12$ are not. A dyadic interval is $[a, b]$ where $a = p/2^k$ and $b = (p+1)/2^k$. For example, [.75, 1] is a dyadic interval but $[1, \pi]$, [0, 2], and [.25, .75] are not. A dyadic cube is the product of dyadic intervals having equal length. The set of dyadic rationals may be denoted as $\mathbb{Q}_2$ and the dyadic lattice as $\mathbb{Q}_2^m$.
    (a) Prove that any two dyadic squares (i.e., planar dyadic cubes) of the same size are either identical or intersect only along their edges.
    (b) Show that the same intersection property is true for dyadic cubes in $\mathbb{R}^m$, when "edge" is interpreted as "$(m-1)$- dimensional face."
22. (a) Given $\epsilon > 0$, show that the unit disc contains finitely many dyadic squares whose total area exceeds $\pi - \epsilon$, and which intersect each other only along their boundaries.
    *(b) Show that the assertion is false for $\epsilon < \pi/2$ if we demand that the dyadic squares be disjoint.
    (c) Formulate (a) in dimension $m = 3$ and $m \geq 4$.
    *(d) Do the analysis with squares and discs interchanged. That is, given $\epsilon > 0$ prove that finitely many disjoint discs can be drawn inside the unit square so that the total area of the discs exceeds $1 - \epsilon$.

*23. Let $b(R)$ and $s(R)$ be the number of integer unit cubes in $\mathbb{R}^m$ that intersect the ball and sphere of radius R, centered at the origin.
    (a) Let $m = 2$ and calculate the limits

$$\lim_{R\to\infty} \frac{s(R)}{b(R)} \quad \text{and} \quad \lim_{R\to\infty} \frac{s(R)^2}{b(R)}.$$

    (b) Take $m \geq 3$. What exponent k makes the limit

$$\lim_{R\to\infty} \frac{s(R)^k}{b(R)}$$

    interesting?
    (c) Let $c(R)$ be the number of integer unit cubes that are contained in the ball of radius R, centered at the origin. Calculate

$$\lim_{R\to\infty} \frac{c(R)}{b(R)}$$

    (d) Shift the ball to a new, arbitrary center (not on the integer lattice) and re-calculate the limits.

*24. Let $f(k, m)$ be the number of k-dimensional faces of the m-cube. See Table 1.

	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$\cdots$	m	$m+1$
$k=0$	2	4	8	$f(0,4)$	$f(0,5)$	$\cdots$	$f(0,m)$	$f(0,m+1)$
$k=1$	1	4	12	$f(1,4)$	$f(1,5)$	$\cdots$	$f(1,m)$	$f(1,m+1)$
$k=2$	0	1	6	$f(2,4)$	$f(2,5)$	$\cdots$	$f(2,m)$	$f(2,m+1)$
$k=3$	0	0	1	$f(3,4)$	$f(3,5)$	$\cdots$	$f(3,m)$	$f(3,m+1)$
$k=4$	0	0	0	$f(4,4)$	$f(4,5)$	$\cdots$	$f(4,m)$	$f(4,m+1)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

Table 1 $f(k, m)$ is the number of k-dimensional faces of the m-cube.

    (a) Verify the numbers in the first three columns.
    (b) Calculate the columns $m = 4$, $m = 5$, and give the formula for passing from the m^{th} column to the $(m+1)^{\text{st}}$.
    (c) What would an $m = 0$ column mean?
    (d) Prove that the alternating sum of the entries in any column is 1. That is, $2 - 1 = 1$, $4 - 4 + 1 = 1$, $8 - 12 + 6 - 1 = 1$, and in general $\sum(-1)^k f(k, m) = 1$. This alternating sum is called the **Euler characteristic**.

25. Prove that the interval $[a, b]$ in $\mathbb{R}$ is the same as the segment $[a, b]$ in $\mathbb{R}^1$. That is,

$$\begin{aligned}&\{x \in \mathbb{R} : a \leq x \leq b\}\\ =&\{y \in \mathbb{R} : \exists\ s, t \in [0,1],\ s + t = 1,\ \text{and } y = sa + tb\}.\end{aligned}$$

    [Hint: How do you prove that two sets are equal?]

26. A **convex combination** of $w_1, \ldots, w_k \in \mathbb{R}^m$ is a vector sum

$$w = s_1 w_1 + \cdots + s_k w_k$$

    such that $s_1 + \cdots + s_k = 1$ and $0 \leq s_1, \ldots, s_k \leq 1$.
    (a) Prove that if a set E is convex then E contains the convex combination of any finite number of points in E.
    (b) Why is the converse obvious?

27. (a) Prove that the ellipsoid

$$E = \{(x, y, z) \in \mathbb{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{x^2} \leq 1\}$$

is convex. [Hint: E is the unit ball for a different dot product. What is it? Does the Cauchy-Schwarz inequality not apply to all dot products?]

(b) Prove that all boxes in $\mathbb{R}^m$ are convex.

28. A function $f : (a, b) \to \mathbb{R}$ is a **convex function** if for all $x, y \in (a, b)$ and all $s, t \in [0, 1]$ with $s + t = 1$,

$$f(sx + ty) \leq sf(x) + tf(y).$$

(a) Prove that f is convex if and only if the set S of points above its graph is convex in $\mathbb{R}^2$. The set S is $\{(x, y) : f(x) \leq y\}$.

*(b) Prove that a convex function is continuous.

(c) Suppose that f is convex and $a < x < u < b$. The slope σ of the line through $(x, f(x))$ and $(u, f(u))$ depends on x and u, $\sigma = \sigma(x, u)$. Prove that σ increases when x increases, and σ increases when u increases.

(d) Suppose that f is second-order differentiable. That is, f is differentiable and its derivative f' is also differentiable. As is standard, write $(f')' = f''$. Prove that f is convex if and only if $f''(x) \geq 0$ for all $x \in (a, b)$.

(e) Formulate a definition of convexity for a function $f : M \to \mathbb{R}$ where $M \subset \mathbb{R}^m$ is a convex set. [Hint: Start with $m = 2$.]

*29. Suppose that E is a convex region in the plane bounded by a curve C.

(a) Show that C has a tangent line except at a countable number of points. [For example, the circle has a tangent line at all its points. The triangle has a tangent line except at three points, and so on.]

(b) Similarly, show that a convex function has a derivative except at a countable set of points.

*30. Suppose that a function $f : [a, b] \to \mathbb{R}$ is monotone increasing: i.e., $x_1 \leq x_2 \Rightarrow f(x_1) \leq f(x_2)$.

(a) Prove that f is continuous except at a countable set of points. [Hint: Show that at each $x \in (a, b)$, f has **right limit** $f(x+)$ and a **left limit** $f(x-)$, which are limits of $f(x+h)$ as h tends to 0 through positive, and negative values respectively. The **jump** of f at x is $f(x+) - f(x-)$. Show that f is continuous at x if and only if it has zero jump at x. At how many points can f have jump ≥ 1? At how many points can the jump be between $1/2$ and 1? Between $1/3$ and $1/2$?]

(b) Is the same assertion true for a monotone function defined on all of $\mathbb{R}$?

31. Find an exact formula for a bijection $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. Is one

$$f(i, j) = j+(1+2+\cdots+(i+j-2)) = \frac{i^2 + j^2 + i(2j - 3) - j + 2}{2}?$$

32. Prove that the union of denumerably many sets B_k, each of which is countable, is countable. How could it happen that the union is finite?

*33. Without using the Schroeder-Bernstein Theorem,

(a) Prove that $[a, b] \sim (a, b] \sim (a, b)$.

(b) More generally, prove that if C is countable then

$$\mathbb{R} \setminus C \quad \sim \quad \mathbb{R} \quad \sim \quad \mathbb{R} \cup C.$$

(c) Infer that the set of irrational numbers has the same cardinality as $\mathbb{R}$, $\mathbb{R} \setminus \mathbb{Q} \sim \mathbb{R}$.

*34. Prove that $\mathbb{R}^2 \sim \mathbb{R}$. [Hint: Think of shuffling two digit strings

$$(a_1a_2a_3\ldots)\&(b_1b_2b_3\ldots) \to (a_1b_1a_2b_2a_3b_3\ldots).$$

In this way you could transform a pair of reals into a single real. Be sure to face the nines-termination issue.]

35. Let S be a set and let $\mathcal{P} = \mathcal{P}(S)$ be the collection of all subsets of S. [$\mathcal{P}(S)$ is called the **power set** of S.] Let $\mathcal{F}$ be the set of functions $f : S \to \{0, 1\}$.

(a) Prove that there is a natural bijection from $\mathcal{F}$ onto $\mathcal{P}$ defined by $f \mapsto \{s \in S : f(s) = 1\}$.

*(b) Prove that the cardinality of $\mathcal{P}$ is greater than the cardinality of S, even when S is empty or finite.

[Hints: The notation Y^X is sometimes used for the set of all functions $X \to Y$. In this notation, $\mathcal{F} = \{0, 1\}^S$ and assertion (b) becomes $\#(S) < \#(\{0, 1\}^S)$. The empty set has one subset, itself, whereas it has no elements, so $\#(\emptyset) = 0$, while $\#(\{0, 1\}^{\emptyset}) = 1$. Assume there is a bijection from S onto $\mathcal{P}$. Then there is a bijection $\beta : S \to \mathcal{F}$, and for each $s \in S$, $\beta(s)$ is a function, say $f_s : S \to \{0, 1\}$. Think like Cantor and try to find a function which corresponds to no s. Infer that β could not have been onto.]

36. A real number is **algebraic** if it is a root of a nonconstant polynomial with integer coefficients.
    (a) Prove that the set A of algebraic numbers is denumerable. [Hint: Each polynomial has how many roots? How many linear polynomials are there? How many quadratics? ...]
    (b) Repeat the exercise for roots of polynomials whose coefficients belong to some fixed, arbitrary denumerable set S.
    *(c) Repeat the exercise for roots of trigonometric polynomials with integer coefficients.
37. A **finite word** is a finite string of letters, say from the roman alphabet.
    (a) What is the cardinality of the set of all finite words, and thus of the set of all possible poems and mathematical proofs?
    (b) What if the alphabet had only two letters?
    (c) What if it had countably many letters?
    (d) Prove that the cardinality of the set Σ_n of all infinite words formed using a finite alphabet of n letters, $n \geq 2$, is equal to the cardinality of $\mathbb{R}$.
    (e) Give a solution to Exercise 34 by justifying the equivalence chain

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} \sim \Sigma_2 \times \Sigma_2 \sim \Sigma_4 \sim \Sigma_2 \sim \mathbb{R}.$$

    (f) How many decimal expansions terminate in an infinite string of 9's? How many don't?
38. If v is a value of a continuous function $f : [a, b] \to \mathbb{R}$ use the l.u.b. property to prove that there are smallest and largest $x \in [a, b]$ such that $f(x) = v$.
39. A function defined on an interval $[a, b]$ or (a, b) is **uniformly continuous** if for each $\epsilon > 0$ there exists a $\delta > 0$ such that $|x - t| < \delta$ implies that $|f(x) - f(t)| < \epsilon$. (Note that this δ can not depend on x, it can only depend on ϵ. With ordinary continuity, the δ can depend on both x and ϵ.)
    (a) Show that a uniformly continuous function is continuous but continuity does not imply uniform continuity. (For example, prove that $\sin(1/x)$ is continuous on the interval $(0, 1)$ but is not uniformly continuous there. Graph it.)
    (b) Is the function $2x$ uniformly continuous on the unbounded interval $(-\infty, \infty)$?
    (c) What about x^2?
*40. Prove that a continuous function defined on an interval $[a, b]$ is uniformly continuous. [Hint: Let $\epsilon > 0$ be given. Think of ϵ as fixed

and consider the sets

$$A(\delta) = \{u \in [a, b] : \text{if } x, t \in [a, u] \text{ and } |x - t| < \delta \text{ then } |f(x) - f(t)| < \epsilon\}$$
$$A = \bigcup_{\delta > 0} A(\delta).$$

Using the least upper bound principle, prove that $b \in A$. Infer that f is uniformly continuous. The fact that continuity on $[a, b]$ implies uniform continuity is one of the important, fundamental principles of continuous functions.]

*41. Define injections $f : \mathbb{N} \to \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{N}$ by $f(n) = 2n$ and $g(n) = 2n$. From f and g, the Schroeder-Bernstein Theorem produces a bijection $\mathbb{N} \to \mathbb{N}$. What is it?
*42. Let (a_n) be a sequence of real numbers. It is **bounded** if the set $A = \{a_1, a_2, \dots\}$ is bounded. The **limit supremum**, or lim sup, of (a_n) as $n \to \infty$ is

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} \left(\sup_{k \geq n} a_k \right)$$

    (a) If the sequence (a_n) is bounded, why does the lim sup exist?
    (b) If $\sup\{a_n\} = \infty$, what is $\limsup a_n$?
    (c) If $\lim_{n\to\infty} a_n = -\infty$, how should we define $\limsup a_n$?
    (d) When is it true that

$$\limsup_{n\to\infty}(a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n$$
$$\limsup_{n\to\infty} ca_n = c \limsup_{n\to\infty} a_n \, ?$$

    (e) Define the **limit infimum**, or lim inf, of a sequence of numbers, and find a formula relating it to the limit supremum.
**43. The unit ball with respect to a norm $\| \quad \|$ on $\mathbb{R}^2$ is

$$\{v \in \mathbb{R}^2 : \|v\| \leq 1\}.$$

    (a) Find necessary and sufficient geometric conditions on a subset of $\mathbb{R}^2$ that it be the unit ball for some norm.
    (b) Give necessary and sufficient geometric conditions that a subset be the unit ball for a norm arising from an inner product.
    (c) Generalize to $\mathbb{R}^m$. [You may find it useful to read about closed sets in the next chapter, and to consult Exercise 38 there.]

44. Assume that V is an inner product space whose inner product induces a norm as $|x| = \sqrt{\langle x, x\rangle}$.
   (a) Show that $|\ |$ obeys the **parallelogram law**

$$|x+y|^2 + |x-y|^2 = 2\,|x|^2 + 2\,|y|^2$$

   for all $x, y \in V$.
   *(b) Show that any norm obeying the parallelogram law arises from a unique inner product. [Hints: Define the prospective inner product as

$$\langle x, y\rangle = \left|\frac{x+y}{2}\right|^2 - \left|\frac{x-y}{2}\right|^2.$$

   Checking that $\langle\ ,\ \rangle$ satisfies the inner product properties of symmetry and positive definiteness is easy. Also it is immediate that $|x|^2 = \langle x, x\rangle$, so $\langle\ ,\ \rangle$ induces the given norm. Checking bilinearity is another story.
      (i) Let $x, y, z \in V$ be arbitrary. Show that the parallelogram law implies

$$\langle x+y, z\rangle + \langle x-y, z\rangle = 2\langle x, y\rangle,$$

      and infer that $\langle 2x, z\rangle = 2\langle x, z\rangle$. For arbitrary $u, v \in V$ set $x = \frac{1}{2}(u+v)$, $y = \frac{1}{2}(u-v)$, plug in to the previous equation, and deduce

$$\langle u, z\rangle + \langle v, z\rangle = \langle u+v, z\rangle,$$

      which is additive bilinearity in the first variable. Why does it now follow at once that $\langle\ ,\ \rangle$ is also additively bilinear in the second variable?
      (ii) To check multiplicative bilinearity, prove by induction that if $m \in \mathbb{Z}$ then $m\langle x, y\rangle = \langle mx, y\rangle$, and if $n \in \mathbb{N}$ then $\frac{1}{n}\langle x, y\rangle = \langle \frac{1}{n}x, y\rangle$. Infer that $r\langle x, y\rangle = \langle rx, y\rangle$ when r is rational. Is $\lambda \mapsto \langle \lambda x, y\rangle - \lambda\langle x, y\rangle$ a continuous function of $\lambda \in \mathbb{R}$, and does this give multiplicative bilinearity?]

2

A Taste of Topology

1 Metric Space Concepts

It may seem paradoxical at first, but a specific math problem can be harder to solve than some abstract generalization of it. For instance, if you want to know how many roots the equation

$$t^5 - 4t^4 + t^3 - t + 1 = 0$$

can have, then you could use calculus and figure it out. It would take a while. But thinking more abstractly, and with less work, you could show that any n^{th} degree polynomial has at most n roots. In the same way many general results about functions of a real variable are more easily grasped at an abstract level — the level of metric spaces.

Metric space theory can be seen as a special case of general topology, and many books present it that way, explaining compactness primarily in terms of open coverings. In my opinion, however, the sequence/subsequence approach provides the easiest and simplest route to mastering the subject. Accordingly it gets top billing throughout this chapter.

A metric space is a set M, the elements of which are referred to as points of M, together with a **metric** d having the three properties that distance has in Euclidean space. The metric $d = d(x, y)$ is a real number defined for all points $x, y \in M$ and $d(x, y)$ is called the distance from the point x to the point y. The three distance properties are: for all $x, y, z \in M$,

(a) **positive definiteness**: $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$.
(b) **symmetry**: $d(x, y) = d(y, x)$.
(c) **triangle inequality**: $d(x, z) \leq d(x, y) + d(y, z)$.

The function d is also called the distance function. Strictly speaking, it is the pair (M, d) which is a metric space, but we will follow the common practice of referring to "the metric space M," and leave to you the job of inferring the correct metric.

The main examples of metric spaces are $\mathbb{R}$, $\mathbb{R}^m$, and their subsets. The metric on $\mathbb{R}$ is $d(x, y) = |x - y|$ where $x, y \in \mathbb{R}$ and $|x - y|$ is the magnitude of $x - y$. The metric on $\mathbb{R}^m$ is the Euclidean length of $x - y$ where x, y are vectors in $\mathbb{R}^m$. Namely,

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_m - y_m)^2}$$

for $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_m)$.

Since Euclidean length satisfies the three distance properties, d is a bona fide metric and it makes $\mathbb{R}^m$ into a metric space. A subset $M \subset \mathbb{R}^m$ becomes a metric space when we declare the distance between points of M to be their Euclidean distance apart as points in $\mathbb{R}^m$. We say that M **inherits** its metric from $\mathbb{R}^m$ and is a **subspace** of $\mathbb{R}^m$. Figure 22 shows a few subsets of $\mathbb{R}^2$ to suggest some interesting metric spaces.

There is also one metric that is hard to picture but valuable as a source for counter-examples, the **discrete metric**. Given any set M define the distance between distinct points to be 1 and the distance between any point and itself to be 0. This is a metric. If M consists of three points, say $M = \{a, b, c\}$, you can think of the vertices of the unit equilateral triangle as a model for M. See Figure 23. They have mutual distance 1 from each other. If M consists of one, two, or four points can you think of a model for the discrete metric on M? More challenging is to imagine the discrete metric on $\mathbb{R}$. All points, by definition of the discrete metric, lie at unit distance from each other.

Convergent sequences and subsequences

A sequence of points in a metric space M is a list $p_1, p_2, \ldots$ where the points p_n belong to M. Repetition is allowed, and not all the points of M need to appear in the list. Good notation for a sequence is (p_n), or $(p_n)_{n\in\mathbb{N}}$. The notation $\{p_n\}$ is also used but it is too easily confused with the set of points making up the sequence. The difference between $(p_n)_{n\in\mathbb{N}}$ and $\{p_n : n \in \mathbb{N}\}$ is that in the former case the sequence prescribes an ordering of the points, while in the latter the points get jumbled together.

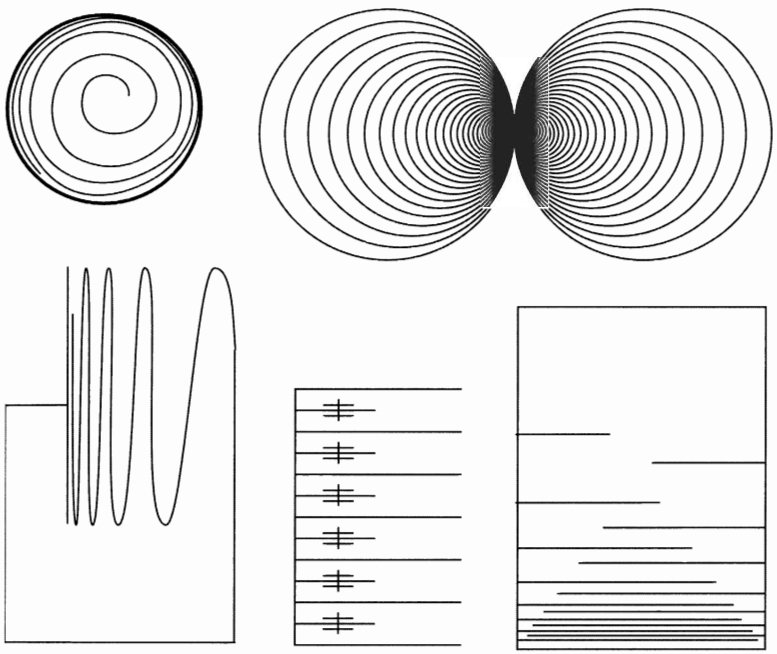


Figure 22 Five metric spaces: a closed outward spiral, a Hawaiian earring, a topologist's sine circle, an infinite television antenna, and Zeno's maze.

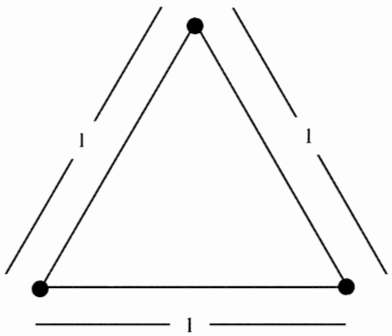


Figure 23 The vertices of the unit equilateral triangle form a discrete metric space.

For example, the sequences $1, 2, 3, \ldots$ and $1, 2, 1, 3, 2, 1, 4, 3, 2, 1, \ldots$ are different sequences but give the same set of points, namely $\mathbb{N}$.

Formally, a sequence is a function $f : \mathbb{N} \to M$. The n^{th} term in the sequence is $f(n) = p_n$. Clearly, every sequence defines a function f :

$\mathbb{N} \to M$ and conversely, every function $f : \mathbb{N} \to M$ defines a sequence in M. The sequence (p_n) **converges to the limit** p in M if

$$\forall \epsilon > 0 \ \exists N \in \mathbb{N} \text{ such that}$$
$$n \in \mathbb{N} \text{ and } n \geq N \quad \Rightarrow \quad d(p_n, p) < \epsilon.$$

Limits are unique in the sense that if (p_n) converges to p and if (p_n) also converges to p' then $p = p'$. The reason is this. Given any $\epsilon > 0$, there are integers N, N' such that if $n \geq N$ then $d(p_n, p) < \epsilon$, while if $n \geq N'$ then $d(p_n, p') < \epsilon$. Then for all $n \geq \max\{N, N'\}$,

$$d(p, p') \leq d(p, p_n) + d(p_n, p') < \epsilon + \epsilon = 2\epsilon.$$

But ϵ is arbitrary and so $d(p, p') = 0$ and $p = p'$. (This is the ϵ-principle again.)

We write $p_n \to p$, or $p_n \to p$ as $n \to \infty$, or

$$\lim_{n \to \infty} p_n = p$$

to indicate convergence. For example, in $\mathbb{R}$ the sequence $p_n = 1/n$ converges to 0 as $n \to \infty$. In $\mathbb{R}^2$ the sequence $(1/n, \sin n)$ does not converge as $n \to \infty$. In the metric space $\mathbb{Q}$ (with the metric it inherits from $\mathbb{R}$) the sequence $1, 1.4, 1.41, 1.414, \ldots$ does not converge.

Just as a set can have a subset, a sequence can have a subsequence. For example, the sequence $2, 4, 6, 8, \ldots$ is a subsequence of $1, 2, 3, 4, \ldots$. The sequence $3, 5, 7, 11, 13, 17, \ldots$ is a subsequence of $1, 3, 5, 7, 9, \ldots$, which in turn is a subsequence of $1, 2, 3, 4, \ldots$. In general, if $(p_n)_{n \in \mathbb{N}}$ and $(q_k)_{k \in \mathbb{N}}$ are sequences and if there is a sequence $1 \leq n_1 < n_2 < n_3 < \ldots$ of integers such that for each $k \in \mathbb{N}$, $q_k = p_{n_k}$ then (q_k) is a **subsequence** of (p_n). Note that the terms in the subsequence occur in the same order as in the mother sequence.

1 Theorem *Every subsequence of a convergent sequence converges and it converges to the same limit as does the mother sequence.*

Proof Let (q_k) be a subsequence of (p_n) $q_k = p_{n_k}$, where $n_1 < n_2 < \ldots$. Assume that (p_n) converges to p in M. Given $\epsilon > 0$, there is an N such that for all $n \geq N$, $d(p_n, p) < \epsilon$. Since $n_1, n_2, \ldots$ are integers, $k \leq n_k$ for all k. Thus, if $k \geq N$ then $n_k \geq N$ and $d(q_k, p) < \epsilon$. Hence (q_k) converges to p. □

A common way to state Theorem 1 is that limits are unaffected when we pass to a subsequence.

Continuity

In linear algebra the objects of interest are linear transformations. In real analysis the objects of interest are functions, especially continuous functions. A function f from the metric space M to the metric space N is just that; $f : M \to N$ and f sends points $p \in M$ to points $fp \in N$. A function is also called a transformation, a mapping, or a map. The function maps M to N. The way you should think of functions — as devices, not formulas — is discussed in Section 4 of Chapter 1.

The most common type of function maps M to $\mathbb{R}$. It is a real-valued function of the variable $p \in M$.

Definition A function $f : M \to N$ is **continuous** if it satisfies the **ϵ, δ condition**:

$$\forall \epsilon > 0 \text{ and } \forall p \in M \ \exists \delta > 0 \text{ such that}$$
$$q \in M \text{ and } d(p, q) < \delta \quad \Rightarrow \quad d(fp, fq) < \epsilon.$$

Here and in what follows, the notation fp is used as convenient shorthand for $f(p)$.

Consider the case of a real valued function of a real variable $f : (a, b) \to \mathbb{R}$ as in Chapter 1. There we defined f to be continuous when

$$\forall \epsilon > 0 \text{ and } \forall x \in (a, b) \ \exists \delta > 0 \text{ such that}$$
$$y \in (a, b) \text{ and } |x - y| < \delta \quad \Rightarrow \quad |fx - fy| < \epsilon.$$

In case $M = (a, b)$ and $N = \mathbb{R}$, the two ϵ, δ definitions are identical except in the second we write out explicitly the distance in $\mathbb{R}$ as $|x - y|$ instead of $d(x, y)$. Thus, every continuous real function of a real variable is an example of a continuous mapping from the metric space (a, b) to the metric space $\mathbb{R}$, and conversely, every continuous map from the metric space (a, b) to the metric space $\mathbb{R}$ is a continuous function of a real variable.

How is continuity expressed using sequences? Nothing could be simpler or more natural.

2 Theorem *$f : M \to N$ is continuous if and only if it sends each convergent sequence in M to a convergent sequence in N, limits being sent to limits.*

Proof Suppose that f is continuous and (p_n) is a convergent sequence in M,

$$\lim_{n \to \infty} p_n = p.$$

Then $(f(p_n))$ is a sequence in N. Continuity implies that given $\epsilon > 0$, there exists $\delta > 0$ such that $d(x, p) < \delta$ implies $d(fx, fp) < \epsilon$. Convergence implies that there is an N such that for all $n \geq N$, $d(p_n, p) < \delta$. Then $d(f(p_n), fp) < \epsilon$ and

$$\lim_{n\to\infty} f(p_n) = fp.$$

We prove the converse in contrapositive form: if f is not continuous then it does not preserve convergence. f being not continuous means that for some $p \in M$ there is an $\epsilon > 0$ such that no matter how small we take δ, there will always be points $x \in M$ with $d(x, p) < \delta$ but $d(fx, fp) \geq \epsilon$. Take

$$\delta_1 = 1,\ \delta_2 = 1/2,\ \ldots\ \delta_n = 1/n,\ \ldots$$

For each δ_n there is a point x_n with $d(x_n, p) < \delta_n = 1/n$ and $d(f(x_n), fp) \geq \epsilon$. Thus

$$\lim_{n\to\infty} x_n = p$$

but $f(x_n)$ does not converge to fp. □

See also Exercise 17.

3 Corollary *The composite of continuous functions is continuous.*

Proof Let $f : M \to N$ and $g : N \to P$ be continuous and assume that

$$\lim_{n\to\infty} p_n = p$$

in M. Since f is continuous, Theorem 2 implies that $\lim_{n\to\infty} f(p_n) = fp$. Since g is continuous, Theorem 2 implies that $\lim_{n\to\infty} g(f(p_n)) = g(fp)$ and therefore $g \circ f : M \to P$ is continuous. See Figure 24. □

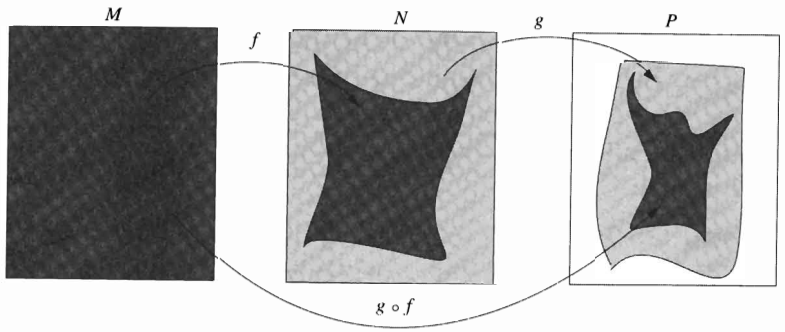


Figure 24 The composite function $g \circ f$.

Moral The sequence condition is the easy way to tell at a glance whether a function is continuous.

Homeomorphism

Vector spaces are isomorphic if there is a linear bijection from one to the other. When are metric spaces isomorphic? They should "look the same." The letters Y and T look the same; and they look different from the letter O. If $f : M \to N$ is a bijection and f is continuous and the inverse bijection $f^{-1} : N \to M$ is also continuous then f is a **homeomorphism**† (or a "homeo" for short) and M, N are homeomorphic. We write $M \cong N$ to indicate that M and N are homeomorphic. $\cong$ is an equivalence relation: $M \cong M$ since the identity map is a homeomorphism $M \to M$; $M \cong N$ clearly implies that $N \cong M$; and the previous corollary shows that the composite of homeomorphisms is a homeomorphism.

Geometrically speaking, a homeomorphism is a bijection that can bend, twist, stretch, and wrinkle the space M to make it coincide with N, but it can not rip, puncture, shred, or pulverize M in the process. The basic questions to ask about metric spaces are:

(a) Given M, N, are they homeomorphic?

(b) What are the continuous functions from M to N?

A major goal of this chapter is to show you how to answer these questions in many cases. For example, is the circle homeomorphic to the interval? To the sphere? etc. Figure 25 indicates that the circle and the (perimeter of the) triangle are homeomorphic, while Figure 14 shows that (a, b), the semicircle, and $\mathbb{R}$ are homeomorphic.

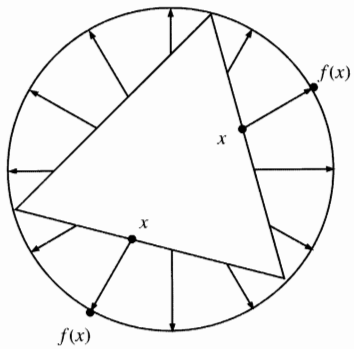


Figure 25 The circle and triangle are homeomorphic.

† This is a rare case in mathematics in which spelling is important. Homeomorphism ≠ homomorphism.

A natural question that should occur to you is whether continuity of f^{-1} is actually implied by continuity of a bijection f. It is not. Here is an instructive example.

Consider the interval $[0, 2\pi) = \{x \in \mathbb{R} : 0 \leq x < 2\pi\}$ and define $f : [0, 2\pi) \to S^1$ to be the mapping $f(x) = (\cos x, \sin x)$ where S^1 is the unit circle in the plane. The mapping f is a continuous bijection, but the inverse bijection is not continuous. For there is a sequence of points (z_n) on S^1 in the fourth quadrant that converges to $p = (1, 0)$, and $f^{-1}(z_n)$ does not converge to $f^{-1}(p) = 0$. Rather it converges to 2π. Thus, f is a continuous bijection whose inverse bijection fails to be continuous. See Figure 26.

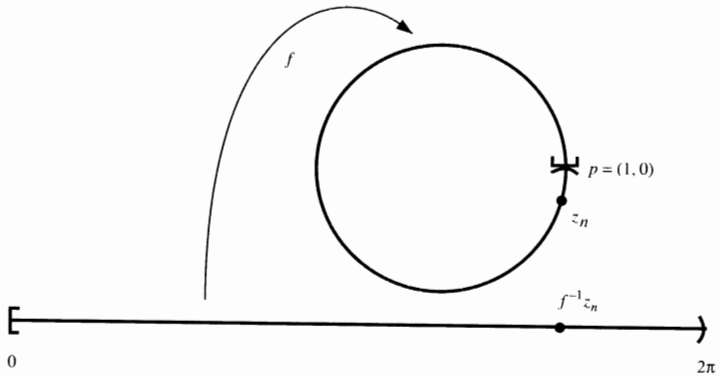


Figure 26 f wraps $[0, 2\pi)$ bijectively onto the circle.

To build your intuition about continuous mappings and homeomorphisms, consider the following examples shown in Figure 27: the unit circle in the plane, a trefoil knot in $\mathbb{R}^3$, the perimeter of a square, the surface of a donut (the 2-torus), the surface of a ceramic coffee cup, the unit interval [0, 1], the unit disc including its boundary. Equip all with the inherited metric. Which should be homeomorphic to which?

Closed Sets and Open Sets

Now we come to two basic concepts in a metric space theory — closedness and openness. Let M be a metric space and S be a subset of M. A point $p \in M$ is a **limit** of S if there exists a sequence (p_n) in S that converges to it.†

† A limit of S is also sometimes called a **limit point** of S. Take care though: some mathematicians require that a limit point of S be the limit of a sequence of *distinct* points of S. They would say that a finite set has no limit points. We will *not* adopt their point of view. Another word used in this context, especially by the French, is "adherence." A point p **adheres** to the set S if and only if p is a limit of S. In more general circumstances, limits are defined using "nets" instead of sequences. You can read more about nets in graduate level topology books such as *Topology* by James Munkres.

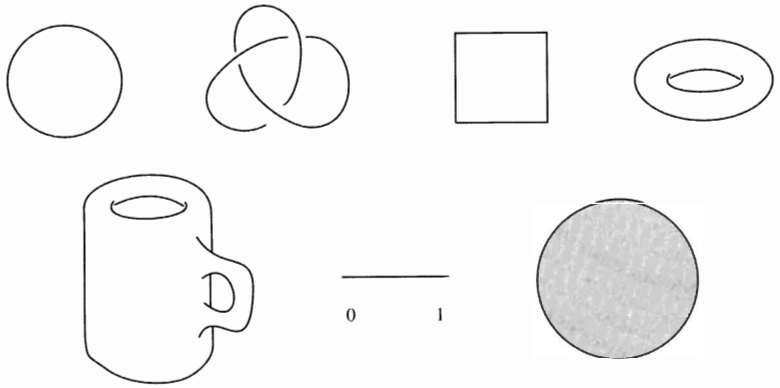


Figure 27 Seven metric spaces.

Definition S is a **closed set** if it contains all its limits.†

Definition S is an **open set** if for each $p \in S$ there exists an $r > 0$ such that

$$d(p, q) < r \quad \Rightarrow \quad q \in S.$$

A little thought will convince you that in the metric space $\mathbb{R}$, the closed interval $[a, b]$ is a closed set and the open interval (a, b) is an open set. See also Theorem 7 below.

4 Theorem *Openness is dual to closedness: the complement of an open set is closed and the complement of a closed set is open.*

Proof Suppose that $S \subset M$ is an open set. We claim that S^c is closed. If $p_n \to p$ and $p_n \in S^c$ we must show that $p \in S^c$. Well, if $p \notin S^c$ then $p \in S$ and, since S is open, there is an $r > 0$ such that

$$d(p, q) < r \quad \Rightarrow \quad q \in S.$$

† Note how similarly algebraists used the word "closed." A field (or group or ring, etc.) is closed under its arithmetic operations: sums, differences, products, and quotients of elements in the field still lie in the field. In our case it is limits. Limits of sequences in S must lie in S.

Since $p_n \to p$, $d(p, p_n) < r$ for all large n, which implies that $p_n \in S$, contrary to the sequence being in S^c. Since the supposition that p lies in S leads to a contradiction, p actually does lie in S^c, proving that S^c is closed.

Suppose that S is closed. We claim that S^c is open. Take any $p \in S^c$. If there fails to exist an $r > 0$ such that

$$d(p, q) < r \quad \Rightarrow \quad q \in S^c,$$

then for each $r = 1/n$, $n = 1, 2, \ldots$, there exists a point $p_n \in S$ such that $d(p, p_n) < 1/n$. This sequence in S converges to $p \in S^c$, contrary to closededness of S. Therefore there actually does exist an $r > 0$ such that

$$d(p, q) < r \quad \Rightarrow \quad q \in S,$$

which proves that S is open. □

Most sets, like doors, are neither open nor closed, but ajar. Keep this in mind. For example, neither $(a, b]$ nor its complement is closed in $\mathbb{R}$; $(a, b]$ is neither closed nor open. Unlike doors, however, sets can be both open and closed at the same time. For example, the empty set $\emptyset$ is a subset of any metric space and it is always closed. There are no sequences and no limits to even worry about. Similarly the full metric space M is a closed subset of itself: it certainly contains the limit of any sequence that converges in M. Thus, $\emptyset$ and M are closed subsets of M. Their complements, M and $\emptyset$, are therefore open: $\emptyset$ and M are both closed and open.

Subsets of M that are both closed and open are **clopen**. See also Exercise 92. It turns out that in $\mathbb{R}$ the only clopen sets are $\emptyset$ and $\mathbb{R}$. In $\mathbb{Q}$, however, things are quite different, sets such as $\{r \in \mathbb{Q} : -\sqrt{2} < r < \sqrt{2}\}$ being clopen in $\mathbb{Q}$. To summarize,

A subset of a metric space can be
closed, open, both, or neither.

You should expect the "typical" subset of a metric space to be neither closed nor open.

The **topology** of M is the collection $\mathcal{T}$ of all open subsets of M.

5 Theorem *$\mathcal{T}$ has three properties:[†] as a system it is closed under union, finite intersection, and it contains $\emptyset$, M. That is,*

[†] Any collection $\mathcal{T}$ of subsets of X that satisfies these three properties is called a topology on X, and X is called a **topological space**. Topological spaces are more general than metric spaces: there exist topologies that do not arise from a metric. Think of them as pathological. The question of which topologies can be generated by a metric and which cannot is discussed in *Topology* by Munkres.

(a) Every union of open sets is an open set.
(b) The intersection of finitely many open sets is an open set.
(c) $\emptyset$ and M are open sets.

Proof (a) If $\{U_\alpha\}$ is any collection of open subsets of M and $V = \bigcup U_\alpha$ then V is open. For if $p \in V$ then p belongs to at least one U_α and there is an $r > 0$ such that

$$d(p, q) < r \quad \Rightarrow \quad q \in U_\alpha.$$

Since $U_\alpha \subset V$, this implies that all such q lie in V, proving that V is open.

(b) If $U_1, \ldots, U_n$ are open sets and $W = \bigcap U_k$ then W is open. For if $p \in W$ then for each k, $1 \le k \le n$, there is an $r_k > 0$ such that

$$d(p, q) < r_k \quad \Rightarrow \quad q \in U_k.$$

Take $r = \min\{r_1, \ldots, r_n\}$. Then

$$d(p, q) < r \quad \Rightarrow \quad q \in U_k,$$

for each k; i.e., $q \in W = \bigcap U_k$, proving that W is open.

(c) It is clear that $\emptyset$ and M are open sets. □

6 Corollary *The intersection of any number of closed sets is a closed set; the finite union of closed sets is a closed set; $\emptyset$ and M are closed sets.*

Proof Take complements and use DeMorgan's Laws. If $\{K_\alpha\}$ is a collection of closed sets then $U_\alpha = (K_\alpha)^c$ is open and

$$K = \bigcap K_\alpha = \left(\bigcup U_\alpha\right)^c.$$

Since $\bigcup U_\alpha$ is open, its complement K is closed. Similarly, a finite union of closed sets is the complement of the finite intersection of their complements, and is a closed set. □

What about an infinite union of closed sets? Generally, it is not closed. For example, the interval $[1/n, 1]$ is closed in $\mathbb{R}$, but the union of these intervals as n ranges over $\mathbb{N}$ is the interval $(0, 1]$ which is not closed. Neither is the infinite intersection of open sets open in general.

Two sets whose closedness/openness properties are basic are:

$$\lim S = \{p \in M : p \text{ is a limit of } S\}$$
$$M_r p = \{q \in M : d(p, q) < r\}$$

The former is the **limit set** of S; the latter is the **r-neighborhood** of p.

7 Theorem lim S *is a closed set and* $M_r p$ *is an open set.*

8 Lemma *The following are equivalent.*
(a) p *is a limit of* S.
(b) $\forall r > 0,\ M_r(p) \cap S \neq \emptyset$.

Proof Assume (a): there is a sequence (p_n) in S with $p_n \to p$. For each $r > 0$ there is an N such that for all $n \geq N$, $p_n \in M_r(p)$, which verifies (b).

Assume (b): each $M_r(p)$ contains a point of S. Take $r = 1, r = 1/2, \ldots, r = 1/n, \ldots$. There exists $p_n \in M_{1/n}(p) \cap S$. The sequence (p_n) converges to p, which verifies (a). □

Proof of Theorem 7 Simple but not immediate! Suppose that $p_n \to p$ and each p_n lies in lim S. Since p_n is a limit of S there is a point $q_n \in S$ such that

$$d(p_n, q_n) < \frac{1}{n}.$$

Then

$$d(p, q_n) \leq d(p, p_n) + d(p_n, q_n) \to 0$$

implies that $q_n \to p$, so $p \in$ lim S and lim S is a closed set.

To check that $M_r p$ is an open set, take any $q \in M_r p$ and observe that

$$s = r - d(p, q) > 0.$$

By the triangle inequality, if $d(q, x) < s$ then

$$d(p, x) \leq d(p, q) + d(q, x) < r,$$

and $M_s q \subset M_r p$. See Figure 28. Since each $q \in M_r p$ has some $M_s q$ that is contained in $M_r p$, $M_r p$ is an open set. □

A **neighborhood** of a point p in M is any open set V that contains p. Theorem 7 implies that $V = M_r p$ is a neighborhood of p. Eventually, you will run across the phrase "closed neighborhood" of p, which refers to a closed set that contains an open set that contains p. Until further notice, all neighborhoods are open.

Usually, sets defined by strict inequalities are open while those defined by equalities or non-strict inequalities are closed. Examples of closed sets in $\mathbb{R}$ are: finite sets, $[a, b]$, $\mathbb{N}$, and the set $\{0\} \cup \{1/n : n \in \mathbb{N}\}$. Each contains all its limits. Examples of open sets in $\mathbb{R}$ are open intervals.

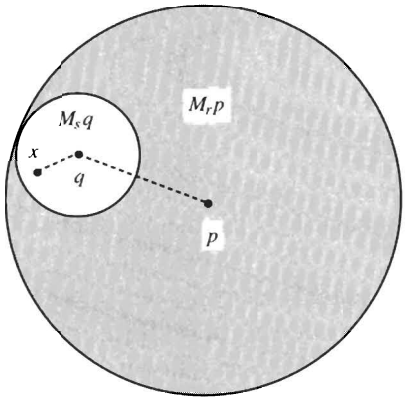


Figure 28 Why the r-neighborhood of p is an open set.

Open subsets of $\mathbb{R}$

Besides bounded open intervals (a, b) with $a, b \in \mathbb{R}$, we consider unbounded open intervals

$$(-\infty, b), \quad (a, \infty), \quad \text{and} \quad \mathbb{R} = (-\infty, \infty).$$

9 Theorem *Every open set* $U \subset \mathbb{R}$ *can be uniquely expressed as a countable union of disjoint open intervals. The endpoints of the intervals do not belong to* U.

Proof For each $x \in U$ define

$$a_x = \inf\{a : \exists\ (a, x) \subset U\}$$
$$b_x = \sup\{b : \exists\ (x, b) \subset U\}.$$

Then $I_x = (a_x, b_x)$ is an open interval, possibly unbounded, and

$$x \in I_x \subset U.$$

It is the *maximal* open interval in U that contains x. If the endpoint b lies in U then, by openness of U, there is some open interval J such that

$$b \in J \subset U.$$

But $I_x \cup J$ is then a larger open interval in U that contains x, contrary to maximality. Thus the endpoints of I_x do not lie in U.

If $y \in U$ is a second point and if $I_x \cap I_y \neq \emptyset$ then $I_x \cup I_y$ is an open interval in U that contains both x and y. By maximality,

$$I_x = I_x \cup I_y = I_y.$$

We conclude that for any $x, y \in U$, the intervals I_x, I_y are either equal or disjoint. This shows that U is the disjoint union of such maximal open intervals.

There can be only countably many disjoint open intervals in $\mathbb{R}$. For, given any collection of disjoint intervals in $\mathbb{R}$, choose a rational number in each. Disjointness implies that the chosen rational numbers are distinct, so there are only countably many of them, and only countably many open intervals. Thus U is the countable disjoint union of the maximal open intervals I_x.

Uniqueness of this division of U into disjoint open intervals is left for the reader to check. □

Topological description of continuity

A property of a metric space or of a mapping between metric spaces that can be described solely in terms of open sets (or equivalently, in terms of closed sets) is called a **topological property**. The next result describes continuity topologically.

Let $f : M \to N$ be given. The **pre-image**[†] of a set $V \subset N$ is

$$f^{\text{pre}}(V) = \{p \in M : f(p) \in V\}.$$

For example, if $f : \mathbb{R}^2 \to \mathbb{R}$ is the function defined by the formula

$$f(x, y) = x^2 + y^2 + 2$$

then the pre-image of the interval $[3, 6]$ in $\mathbb{R}$ is the annulus in the plane with inner radius 1 and outer radius 2. Figure 29 shows the domain of f as $\mathbb{R}^2$ and the target as $\mathbb{R}$. The range is the set of real numbers ≥ 2. The graph of f is a paraboloid with lowest point $(0, 0, 2)$. The second part of the figure shows the portion of the graph lying above the annulus. You will find it useful to keep in mind the distinctions among the concepts: function, range, and graph.

[†] The pre-image of V is also called the **inverse image** of V and is denoted by $f^{-1}(V)$. Unless f is a bijection, this notation leads to confusion. There is no map f^{-1} and yet expressions like $V \supset f(f^{-1}(V))$ are written that mix maps and non-maps. By the way, if f sends no point of M into V then $f^{\text{pre}}(V)$ is the empty set.

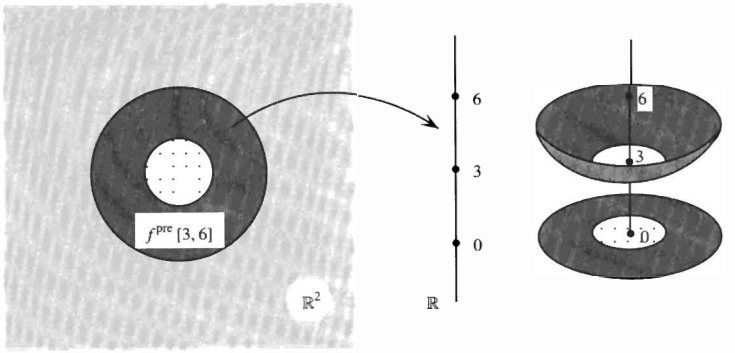


Figure 29 The function $f : (x, y) \mapsto x^2 + y^2 + 2$ and its graph over the pre-image of $[3, 6]$.

10 Theorem *The following are equivalent for continuity of $f : M \to N$.*

(i) The ϵ, δ condition.

*(ii) The **closed set condition**: the pre-image of each closed set in N is closed in M.*

*(iii) The **open set condition**: the pre-image of each open set in N is open in M.*

Proof Totally natural! Assume (i). Let K be closed in N and let (p_n) be a sequence in $f^{\text{pre}}(K)$ that converges to p in M. We must show that $p \in f^{\text{pre}}(K)$. According to Theorem 2, f preserves limits so

$$\lim_{n\to\infty} f(p_n) = fp.$$

By assumption $f(p_n) \in K$. Since K is closed, $fp \in K$. Hence $p \in f^{\text{pre}}(K)$, which gives (ii).

Since $(f^{\text{pre}}(U))^c = f^{\text{pre}}(U^c)$, (ii) $\Rightarrow$ (iii).

Assume (iii). Let $p \in M$ and $\epsilon > 0$ be given. According to Theorem 7, $N_\epsilon(fp)$ is open in N. By (iii), $f^{\text{pre}}(N_\epsilon(fp))$ is open in M, and p belongs to it, so there is some $M_\delta(p) \subset f^{\text{pre}}(N_\epsilon(fp))$. Thus,

$$d(x, p) < \delta \Rightarrow d(fx, fp) < \epsilon,$$

and f is continuous. See Figure 30. □

I hope that you find the closed and open set characterizations of continuity elegant. Note that no explicit mention is made of the metric. The open set condition is purely topological. It would be perfectly valid to take as a *definition* of continuity that the pre-image of each open set is open. In fact this is exactly what's done in general topology.

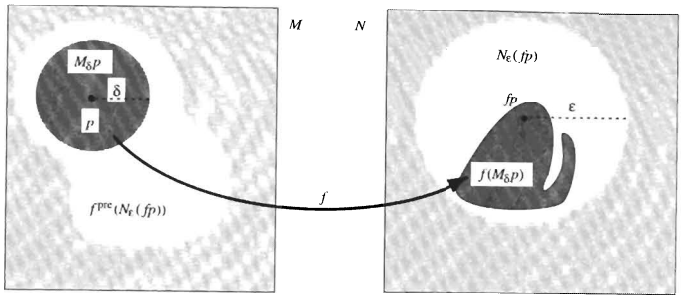


Figure 30 The ϵ, δ - condition for a continuous function $f : M \to N$.

11 Corollary *A homeomorphism $f : M \to N$ bijects the collection of open sets in M to the collection of open sets in N. It bijects the topologies.*

Proof Let V be an open set in N. By Theorem 10, since f is continuous, the pre-image of V is open in M. Since f is a bijection, this pre-image $U = \{p \in M : f(p) \in V\}$ is exactly the image of V by the inverse bijection, $U = f^{-1}(V)$. The same thing can be said about f^{-1} since f^{-1} is also a homeomorphism. That is, $V = f(U)$. Thus, sending U to $f(U)$ bijects the topology of M to the topology of N. □

Because of this corollary, a homeomorphism is also called a **topological equivalence**.

In general, continuous maps do not need to send open sets to open sets. For example, the squaring map $x \mapsto x^2$ from $\mathbb{R}$ to $\mathbb{R}$ sends the open interval $(-1, 1)$ to the non-open interval $[0, 1)$. See also Exercise 34.

Closure, Interior, and Boundary

Let S be a subset of the metric space M. Its closure, interior, and boundary are defined as follows:

closure $\overline{S} = \bigcap K$ where K ranges through the collection of all closed sets that contain S. Equivalently,

$$\overline{S} = \{x \in M : \text{ if } K \text{ is closed and } S \subset K \text{ then } x \in K\}.$$

interior $\text{int}(S) = \bigcup U$ where U ranges through the collection of all open sets contained in S. Equivalently,

$$\text{int } S = \{x \in M : \text{ for some open } U \subset S, x \in U\}.$$

boundary $\partial S = \overline{S} \cap \overline{S^c}$.

Since the intersection of closeds is closed, $\overline{S}$ and ∂S are closed. By construction, $\overline{S}$ is the smallest closed set that contains S because it is contained in every closed set that contains S. Similarly the interior of S is the largest open set contained in S, and the boundary is the closure minus the interior.

Alternate notations for closure, interior, and boundary are

$$\begin{aligned} \overline{S} &= \text{cl}(S) \\ \text{int}(S) &= \mathring{S} \\ \partial S &= \dot{S} = \text{bd}(S) = \text{fr}(S). \end{aligned}$$

The last stands for "frontier." Several useful facts about closure, interior, and boundary are presented in the exercises. See Exercise 92 in particular.

12 Proposition $\overline{S} = \lim S$.

Proof By Theorem 7, $\lim S$ is a closed set. It contains S, because for each $s \in S$, the constant sequence $(s, s, \dots)$ converges to s. By minimality of the closure,

$$\overline{S} \subset \lim S.$$

On the other hand, $\lim S$ consists of limits of S, and because $\overline{S}$ is closed it must contain these limits, so

$$\lim S \subset \overline{S},$$

and the two sets are equal. □

Inheritance

Suppose that S is a subset of the metric space M and also of the metric space N. If S is open or closed in M, is the same true when S is judged in N? Not necessarily. The interval (a, b) is closed in itself, as is any metric space, but it is not closed in $\mathbb{R}$. It fails to contain its limits a, b in $\mathbb{R}$.

Even more striking is the situation concerning the set

$$S = \{x \in \mathbb{Q} : -\sqrt{2} < x < \sqrt{2}\}.$$

As a subset of $\mathbb{Q}$, S is clopen (both closed and open) but as a subset of $\mathbb{R}$ it is neither closed nor open. It contains all its limits in $\mathbb{Q}$ but not all its limits in $\mathbb{R}$, and the same is true of its complement $\mathbb{Q} \setminus S$.

The notation used above for closure, $\overline{S}$, does not take into account the ambient metric space M. To clarify the metric space / metric subspace

situation we can write $\mathrm{cl}_M(S)$ to indicate the closure of S relative to the metric space M. It is the set of all limits in M of S. Similarly we can write $\partial_M(S)$ and $\mathrm{int}_M(S)$ to indicate that we are working in M.

Let N be a metric subspace of M. The r-neighborhood of $p \in N$, $N_r p$, is the set of points in N whose distance to p is less than r. It is exactly the same as $M_r p \cap N$. Closures behave the same way. If $S \subset N$ then $\mathrm{cl}_N(S) = \mathrm{cl}_M(S) \cap N$. See Figure 31.

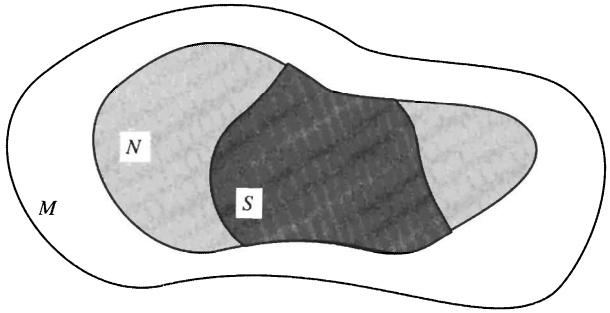


Figure 31 The closure of S in N is its closure in M intersected with N.

13 Inheritance Principle *If $K \subset N \subset M$ where M is a metric space and N is a metric subspace then K is closed in N if and only if there is some subset L of M such that L is closed in M and $K = L \cap N$. That is, N **inherits** its closeds from M.*

Proof Assume that K is closed in N, and consider the set $L = \mathrm{cl}_M(K)$. It is a closed subset of M and consists of K together with its limits in M. See Figure 32.

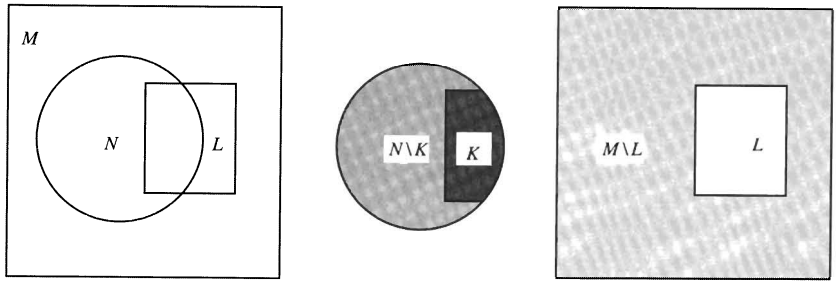


Figure 32 The relations among K, L, M, N.

None of these limits in M lies in $N \setminus K$ since K is closed in N. Hence $K = L \cap N$. On the other hand if L is any closed set in M then $L \cap N$ contains all its limits in N and $L \cap N$ is closed in N. □

14 Corollary *Dually, a metric subspace inherits its opens from the big metric space.*

Proof Take complements. □

15 Corollary *Assume that N is a metric subspace of M and also is a closed subset of M. A set $K \subset N$ is closed in N if and only if it is closed in M.*

Proof Assume that K is closed in N. If p is a limit of K then p is also a limit of N and lies in N since N is closed in M. Since K is closed in N, p also lies in K. Thus K is closed in M. On the other hand if $K \subset N$ is closed in M then it contains all its limits in M and thus contains all its limits in N. □

16 Corollary *Assume that N is a metric subspace of M and also is an open subset of M. A set $U \subset N$ is open in N if and only if it is open in M.*

Proof The proof is left to the reader as Exercise 25. □

Clustering and Condensing

Two concepts similar to limits are often used in metric spaces — clustering and condensing. The set S **clusters** at p (and p is a **cluster point**[†] of S) if each $M_r p$ contains infinitely many points of S. The set S **condenses** at p (and p is a **condensation point** of S) if each $M_r p$ contains uncountably many points of S. Thus, S limits at p, clusters at p, or condenses at p according to whether each $M_r p$ contains some, infinitely many, or uncountably many points of S. See Figure 33.

17 Theorem *The following are equivalent conditions to S clustering at p.*

(i) There is a sequence of distinct points in S that converges to p.
(ii) Each neighborhood of p contains infinitely many points of S.
(iii) Each neighborhood of p contains at least two points of S.
(iv) Each neighborhood of p contains at least one point of S other than p.

Proof Clearly (i) ⇒ (ii) ⇒ (iii) ⇒ (iv), and (ii) is the definition of clustering. It remains to check (iv) ⇒ (i).

[†] Cluster points are also called **accumulation points**. As mentioned above, they are also sometimes called limit points, a usage that conflicts with the limit idea. A finite set S has no cluster points, but of course, each of its points p is a limit of S since the constant sequence $(p, p, p, \ldots)$ converges to p.

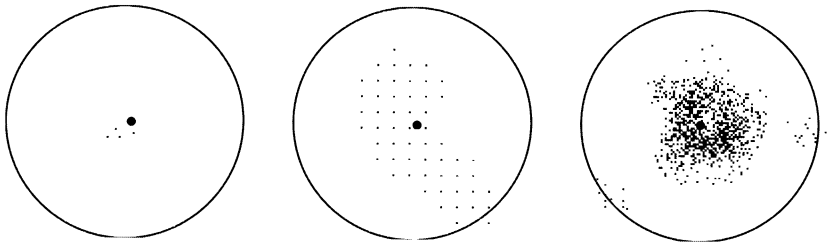


Figure 33 Limiting, clustering, and condensing behavior.

Assume (iv) is true: each neighborhood of p contains a point of S other than p. In $M_1 p$ choose a point $p_1 \in (S \setminus \{p\})$. Set $r_2 = \min(1/2, d(p_1, p))$, and in the smaller neighborhood $M_{r_2} p$, choose $p_2 \in (S \setminus \{p\})$. Proceed inductively: set $r_n = \min(1/n, d(p_{n-1}, p))$ and in $M_{r_n} p$, choose $p_{n+1} \in (S \setminus \{p\})$. The points p_n are distinct since they have different distances to p,

$$d(p_1, p) \geq r_2 > d(p_2, p) \geq r_3 > d(p_3, p) \geq \ldots.$$

Thus (iv) $\Rightarrow$ (i) and the four conditions are equivalent. See Figure 34. □

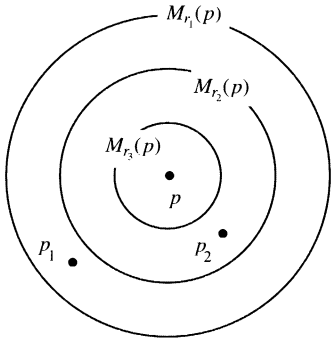


Figure 34 The sequence of distinct points that converge to p.

Condition (iv) is the form of the definition of clustering most frequently used, although it is the hardest to grasp. It is customary to denote by S' the set of cluster points of S.

18 Proposition $S \cup S' = \overline{S}$.

Proof A cluster point is a type of limit of S, so $S' \subset \lim S$ and

$$S \cup S' \subset \lim S.$$

On the other hand, if $p \in \lim S$ then either $p \in S$ or else each neighborhood of p contains points of S other than p. This implies that $p \in S \cup S'$,

$$\lim S \subset S \cup S',$$

and the two sets are equal. □

19 Corollary *S is closed if and only if $S' \subset S$.*

Proof S is closed if and only if $S = \overline{S}$. Since $\overline{S} = S \cup S'$, equivalent to $S \subset S'$ is $\overline{S} = S$. □

20 Corollary *The least upper bound and greatest lower bound of a nonempty bounded subset $S \subset \mathbb{R}$ belong to the closure of S. Thus, if S is closed, they belong to S.*

Proof If $b =$ l.u.b. S then each interval $(b - r, b]$ contains points of S. The same is true for intervals $[a, a + r)$ where $a =$ g.l.b. S. □

Product Metrics

We next define a metric on the Cartesian product $M = M_1 \times M_2$ of two metric spaces. There are three natural ways to do so:

$$d_E(p, q) = \sqrt{d_1(p_1, q_1)^2 + d_2(p_2, q_2)^2}$$
$$d_{\max}(p, q) = \max\{d_1(p_1, q_1),\ d_2(p_2, q_2)\}$$
$$d_{\text{sum}}(p, q) = d_1(p_1, q_1) + d_2(p_2, q_2)$$

where $p = (p_1, p_2)$ and $q = (q_1, q_2)$ belong to M. (d_E is the Euclidean product metric.) The proof that these expressions actually define metrics on M is left as Exercise 83.

21 Theorem (Convergence in a product space) *The following are equivalent for a sequence $p_n = (p_{1n}, p_{2n})$ in $M = M_1 \times M_2$:*

(a) (p_n) converges with respect to the metric $d_{\max}$.
(b) (p_n) converges with respect to the metric d_E.
(c) (p_n) converges with respect to the metric d_{sum}.
(d) (p_{1n}) and (p_{2n}) converge in M_1 and M_2 respectively.

Proof The issue is **comparability** of the metrics. We claim that

$$d_{\max} \leq d_E \leq d_{\text{sum}} \leq 2d_{\max}.$$

Dropping the smaller term inside the square root shows that $d_{\max} \le d_E$; comparing the square of d_E and the square of d_{sum} shows that the latter has the terms of the former and the cross term besides, so $d_E \le d_{\text{sum}}$; and clearly d_{sum} is no larger than twice its greater term, so $d_{\text{sum}} \le 2d_{\max}$. From comparability, equivalence of (a) – (d) is immediate. □

22 Corollary *A sequence in a Cartesian product of m metric spaces converges with respect to the sum metric if and only if it converges with respect to the max metric if and only if each component sequence converges.*

Proof $d_{\max} \le d_E \le d_{\text{sum}} \le m\, d_{\max}$. □

23 Corollary (Convergence in $\mathbb{R}^m$) *A sequence of vectors (v_n) in $\mathbb{R}^m$ converges in $\mathbb{R}^m$ if and only if each of its component sequences (v_{in}) converges, $1 \le i \le m$. The limit of the vector sequence is the vector*

$$v = \lim_{n\to\infty} v_n = \left(\lim_{n\to\infty} v_{1n}, \lim_{n\to\infty} v_{2n}, \ldots, \lim_{n\to\infty} v_{mn}\right).$$

Proof This a special case of the previous corollary. □

Use of $d_{\max}$ and d_{sum} often simplifies proofs by avoiding square root manipulations. The sum metric is also called the **Manhattan metric** or the **taxicab metric**. Figure 35 shows the "unit discs" with respect to these metrics in $\mathbb{R}^2$.

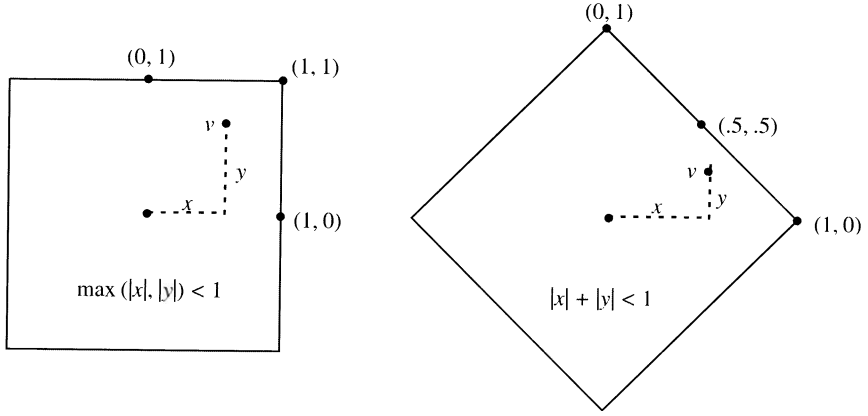


Figure 35 The unit disc in the max metric is a square, and in the sum metric it is a rhombus.

Continuity of Arithmetic in $\mathbb{R}$

Addition is a mapping $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ that assigns to (x, y) the real number $x + y$. Subtraction and multiplication are also such mappings. Division is a mapping $\mathbb{R} \times (\mathbb{R} \setminus \{0\}) \to \mathbb{R}$ that assigns to (x, y) the number x/y.

24 Theorem *The arithmetic operations of $\mathbb{R}$ are continuous.*

Proof Let $(x_0, y_0) \in \mathbb{R} \times \mathbb{R}$ be given and let $\epsilon > 0$ be given. Set $s_0 = |x_0| + |y_0|$.

(+) Take $\delta = \epsilon$. If $|x - x_0| + |y - y_0| < \delta$ then

$$|(x + y) - (x_0 + y_0)| \le |x - x_0| + |y - y_0| < \delta = \epsilon.$$

(−) Take $\delta = \epsilon$. If $|x - x_0| + |y - y_0| < \delta$ then

$$|(x - y) - (x_0 - y_0)| \le |x - x_0| + |y - y_0| < \delta = \epsilon.$$

(×) Take $\delta = \min(1,\ \epsilon/(1 + s_0))$. If $|x - x_0| + |y - y_0| < \delta$ then

$$|(x \cdot y) - (x_0 \cdot y_0)| \le |x|\,|y - y_0| + |x - x_0|\,|y_0| < \epsilon.$$

(÷) Take $\delta = \min(|y_0|/2,\ 1,\ \epsilon y_0^2/(2s_0 + 2))$. If $|x - x_0| + |y - y_0| < \delta$ then

$$|(x \div y) - (x_0 \div y_0)| \le \frac{|xy_0 - xy| + |xy - x_0y|}{|yy_0|} < \epsilon.$$

25 Corollary *The sums, differences, products, and quotients of real valued continuous functions are continuous. (The denominator functions should not equal zero.)*

Proof Take, for example, the sum $f + g$ where $f, g : M \to \mathbb{R}$ are continuous. It is the composite of continuous functions

$$M \xrightarrow{f\times g} \mathbb{R}\times\mathbb{R} \xrightarrow{+} \mathbb{R}$$
$$x \mapsto (f(x), g(x)) \mapsto \text{Sum}(f(x), g(x)),$$

and is therefore continuous. The same applies to the other operations. □

Cauchy Sequences

In Chapter 1 we discussed the Cauchy criterion for convergence of a sequence of real numbers. There is a natural way to carry these ideas over to a metric space M. The sequence (p_n) in M satisfies a **Cauchy condition** provided that for each $\epsilon > 0$ there is an integer N such that for all $k, n \ge N$, $d(p_k, p_n) < \epsilon$, and (p_n) is said to be a **Cauchy sequence**. In symbols,

$$\forall \epsilon > 0 \ \exists N \text{ such that } k, n \ge N \Rightarrow d(p_k, p_n) < \epsilon.$$

The terms of a Cauchy sequence "bunch together" as $n \to \infty$. Each convergent sequence (p_n) is Cauchy. For if (p_n) converges to p as $n \to \infty$ then, given $\epsilon > 0$, there is an N such that for all $n \geq N$,

$$d(p_n, p) < \frac{\epsilon}{2}.$$

By the triangle inequality, if $k, n \geq N$ then

$$d(p_k, p_n) \leq d(p_k, p) + d(p, p_n) < \epsilon,$$

so convergence $\Rightarrow$ Cauchy.

Theorem 1.5 states that the converse is true in the metric space $\mathbb{R}$. Every Cauchy sequence in $\mathbb{R}$ converges to a limit in $\mathbb{R}$. In the general metric space, however, this need not be true. For example, consider the metric space $\mathbb{Q}$ of rational numbers, equipped with the standard distance $d(x, y) = |x - y|$, and consider the sequence

$$(r_n) = (1.4, 1.41, 1.414, 1.4142, \ldots).$$

It is Cauchy. Given $\epsilon > 0$, choose $N > -\log_{10} \epsilon$. If $k, n \geq N$ then $|r_k - r_n| \leq 10^{-N} < \epsilon$. Nevertheless, (r_n) refuses to converge in $\mathbb{Q}$. After all, as a sequence in $\mathbb{R}$ it converges to $\sqrt{2}$, and if it also converges to some $r \in \mathbb{Q}$, then by uniqueness of limits in $\mathbb{R}$, $r = \sqrt{2}$, something we know is false. In brief, convergence $\Rightarrow$ Cauchy but not conversely.

A metric space M is **complete** if each Cauchy sequence in M converges to a limit in M. Theorem 1.5 implies that $\mathbb{R}$ is complete.

26 Theorem *$\mathbb{R}^m$ is complete.*

Proof Let (p_n) be a Cauchy sequence in $\mathbb{R}^m$. Express p_n in components as

$$p_n = (p_{1n}, \ldots, p_{mn}).$$

Because (p_n) is Cauchy, each component sequence $(p_{jn})_{n \in \mathbb{N}}$ is Cauchy. Completeness of $\mathbb{R}$ implies that the component sequences converge, and therefore the vector sequence converges. □

27 Theorem *Every closed subset of a complete metric space is a complete metric space.*

Proof Let A be a closed subset of the complete metric space M and let (p_n) be a Cauchy sequence in A. It is of course also a Cauchy sequence in M and therefore it converges to a limit p in M. Since A is closed, $p \in A$. □

28 Corollary *Every closed subset of Euclidean space is a complete metric space.*

Proof Obvious from the previous theorem and completeness of $\mathbb{R}^m$. □

Boundedness

A subset S of a metric space M is **bounded** if for some $p \in M$ and some $r > 0$,

$$S \subset M_r p.$$

A set which is not bounded is **unbounded**. For example, the elliptical region $4x^2 + y^2 < 4$ is a bounded subset of $\mathbb{R}^2$, while the hyperbola $xy = 1$ is unbounded.

Distinguish the word "bounded" from the word "finite." The first refers to physical size, the second to the number of elements. The concepts are totally different. Also, boundedness has little connection to the existence of a boundary — a clopen subset of a metric space has empty boundary, but some clopen sets are bounded.

It is easy to check that the terms of a Cauchy sequence (p_n) in M form a bounded subset of M. Simply take $\epsilon = 1$, and apply the Cauchy definition: there is an N such that for all $n, m \geq N$, $d(p_1, p_m) < 1$. Choose

$$r > 1 + \max\{d(p_1, p_2), \ldots, d(p_1, p_N)\}.$$

By the triangle inequality, the Cauchy sequence lies entirely in $M_r(p_1)$. Since a convergent sequence is Cauchy, its terms also form a bounded set.

Boundedness is not a topological property. For example, consider $(-1, 1)$ and $\mathbb{R}$. They are homeomorphic, although $(-1, 1)$ is bounded and $\mathbb{R}$ is unbounded. The same example shows that completeness is not a topological property.

A function from M to another metric space N is a **bounded function** if its range is bounded. That is, there exist $q \in N$ and $r > 0$ such that

$$fM \subset N_r q.$$

Note that a function can be bounded even though its graph is not. For example, $x \mapsto \sin x$ is a bounded function $\mathbb{R} \to \mathbb{R}$ although its graph, $\{(x, y) \in \mathbb{R}^2 : y = \sin x\}$, is an unbounded subset of $\mathbb{R}^2$.

2 Compactness

Compactness is the single most important concept in real analysis. It is what reduces the infinite to the finite.

A subset A of a metric space M is (sequentially) **compact** if every sequence (a_n) in A has a subsequence (a_{n_k}) that converges to a limit in A.

The empty set and finite sets are trivial examples of compact sets. For a sequence (a_n) contained in a finite set repeats a term infinitely often, and the corresponding constant subsequence converges.

Compactness is a *good* feature of a set. We will develop criteria to decide whether a set is compact. The first is the most often used, but beware! — its converse is generally false.

29 Theorem *Every compact set is closed and bounded.*

Proof Suppose that A is a compact subset of the metric space M and that p is a limit of A. There is a sequence (a_n) in A converging to p. By compactness, some subsequence (a_{n_k}) converges to some $q \in A$, but every subsequence of a convergent sequence converges to the same limit as does the mother sequence, so $q = p$ and $p \in A$. Thus, A is closed.

To see that A is bounded, choose and fix any point $p \in M$. Either A is bounded or else for each $n \in \mathbb{N}$ there is a point $a_n \in A$ such that $d(p, a_n) \geq n$. Compactness implies that some subsequence (a_{n_k}) converges. Convergent sequences are bounded, which contradicts the fact that $d(p, a_{n_k}) \to \infty$ as $k \to \infty$. Therefore (a_n) can not exist and A is bounded. □

30 Theorem *The closed interval $[a, b] \subset \mathbb{R}$ is compact.*

Proof Let (x_n) be a sequence in $[a, b]$. Let

$$C = \{x \in [a, b] : x_n < x \text{ only finitely often}\}.$$

Since $a \in C$, $C \neq \emptyset$. Clearly b is an upper bound for C. By the least upper bound property of $\mathbb{R}$, $c = \text{l.u.b.}\, C$ exists, $c \in [a, b]$. We claim that a subsequence of (x_n) converges to c. Suppose not, i.e., no subsequence of (x_n) converges to c. Then for some $\epsilon > 0$, x_n lies in $(c - \epsilon, c + \epsilon)$ only finitely often, which implies that $c + \epsilon \in C$, contrary to c being an upper bound for C. Hence some subsequence of (x_n) does converge to c, and $[a, b]$ is compact. □

To pass from $\mathbb{R}$ to $\mathbb{R}^m$, think about compactness in terms of Cartesian products.

31 Theorem *The Cartesian product of two compact sets is compact.*

Proof Let $(a_n, b_n) \in A \times B$ be given where $A \subset M$ and $B \subset N$ are compact. There exists a subsequence a_{n_k} that converges to some point $a \in A$ as $k \to \infty$. The subsequence (b_{n_k}) has a sub-subsequence $(b_{n_{k(\ell)}})$ that converges to some $b \in B$ as $\ell \to \infty$. The sub-subsequence $(a_{n_{k(\ell)}})$ continues to converge to the point a. Thus

$$(a_{n_{k(\ell)}}, b_{n_{k(\ell)}}) \to (a, b)$$

as $\ell \to \infty$. This implies that $A \times B$ is compact. □

32 Corollary *The Cartesian product of m compact sets is compact.*

Proof Write $A_1 \times A_2 \times \cdots \times A_m = A_1 \times (A_2 \times \cdots \times A_m)$ and perform induction on m. (Theorem 31 handles the bottom case $m = 2$.) □

33 Corollary *A box $[a_1, b_1] \times \cdots \times [a_m, b_m]$ is compact.*

Proof Obvious from Theorem 30 and the previous corollary. □

An equivalent formulation of these results is the

34 Bolzano-Weierstrass Theorem *Any bounded sequence in $\mathbb{R}^m$ has a convergent subsequence.*

Proof A bounded sequence is contained in a box, which is compact, and therefore the sequence has a subsequence that converges to a limit in the box. □

Here is a simple fact about compacts.

35 Theorem *A closed subset of a compact set is compact.*

Proof If A is a closed subset of the compact set C and if (a_n) is a sequence of points in A, then clearly (a_n) is also a sequence of points in C, so by compactness of C, there is a subsequence (a_{n_k}) converging to a limit $p \in C$. Since A is closed, p lies in A, which proves that A is compact. □

Now we come to the first partial converse to Theorem 29.

36 Heine-Borel Theorem *Every closed and bounded subset of $\mathbb{R}^m$ is compact.*

Proof Let $A \subset \mathbb{R}^m$ be closed and bounded. Boundedness implies that A is contained in some box, which is compact. Since A is closed, Theorem 35 implies that A is compact. □

The Heine-Borel Theorem states that closed and bounded subsets of Euclidean space are compact, but it is *vital* to remember that a closed and bounded subset of a general metric space may fail to be compact. For example, the set S of rational numbers in $[0, 1]$ is a closed, bounded, noncompact subset of $\mathbb{Q}$. The set S is noncompact because it contains sequences which have no subsequences that converge in S.

Ten Examples of Compact Sets

1. Any finite subset of a metric space; for instance, the empty set.
2. Any closed subset of a compact set.
3. The union of finitely many compact sets.
4. The Cartesian product of finitely many compact sets.
5. The intersection of arbitrarily many compact sets.
6. The unit ball in $\mathbb{R}^3$.
7. The boundary of a compact set, for instance the unit 2-sphere in $\mathbb{R}^3$.
8. The set $\{x \in \mathbb{R} : \ \exists n \in \mathbb{N} \text{ and } x = 1/n\} \cup \{0\}$.
9. The Hawaiian earring. See page 53.
10. The Cantor set. See Section 5.

Nests of Compacts

If $A_1 \supset A_2 \supset \cdots \supset A_n \supset A_{n+1} \supset \ldots$ then (A_n) is a **nested sequence** of sets. Its intersection is

$$\bigcap A_n = \{p : \text{ for each } n, \quad p \in A_n\}.$$

See Figure 36.

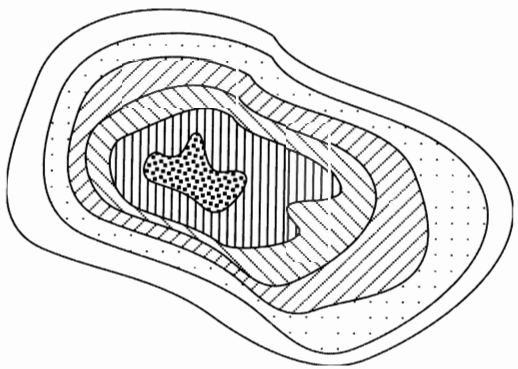


Figure 36 A nested sequence of sets.

For example, we could take A_n to be the disc $\{z \in \mathbb{R}^2 : |z| \leq 1/n\}$. The intersection of all the sets A_n is then the singleton $\{0\}$. On the other hand, if A_n is the ball $\{z \in \mathbb{R}^3 : |z| \leq 1 + 1/n\}$ then $\bigcap A_n$ is the closed unit ball B^3.

37 Theorem *The intersection of a nested sequence of compact non-empty sets is compact and non-empty.*

Proof Let (A_n) be such a sequence. By Theorem 29, A_n is closed. The intersection of closed sets is always closed. Thus, $\bigcap A_n$ is a closed subset of the compact set A_1, and is therefore compact. It remains to show that the intersection is non-empty.

A_n is non-empty, so for each $n \in \mathbb{N}$ we can choose $a_n \in A_n$. The sequence (a_n) lies in A_1 since the sets are nested. Compactness of A_1 implies that (a_n) has a subsequence (a_{n_k}) converging to some point $p \in A_1$. The limit p also lies in the set A_2 since except possibly for the first term, the subsequence (a_{n_k}) lies in A_2 and A_2 is a closed set. The same is true for A_3 and for all the sets in the nested sequence. Thus, $p \in \bigcap A_n$ and $\bigcap A_n$ is shown to be non-empty. □

The **diameter** of a non-empty set $S \subset M$ is the supremum of the distances $d(x, y)$ between points of S.

38 Corollary *If in addition to being nested, non-empty, and compact, the sets A_n have diameter that tends to 0 as $n \to \infty$, then $A = \bigcap A_n$ is a single point.*

Proof For each $k \in \mathbb{N}$, A is a subset of A_k, which implies that A has diameter zero. Since any distinct points lie at positive distance from each other, A consists of at most one point, while by Theorem 37 it consists of at least one point. See also Exercise 27. □

A nested sequence of non-empty *noncompact* sets can have empty intersection. For example, the open discs with center $(1/n, 0)$ on the x-axis and radius $1/n$ are nested, $n \in \mathbb{N}$, but contain no common point. (Their closures do intersect at a common point, the origin.) See Figure 37.

Continuity and Compactness

Next we discuss how compact sets behave under continuous transformations.

39 Theorem *If $f : M \to N$ is continuous and A is a compact subset of M then fA is a compact subset of N. That is, the continuous image of a compact is compact.*

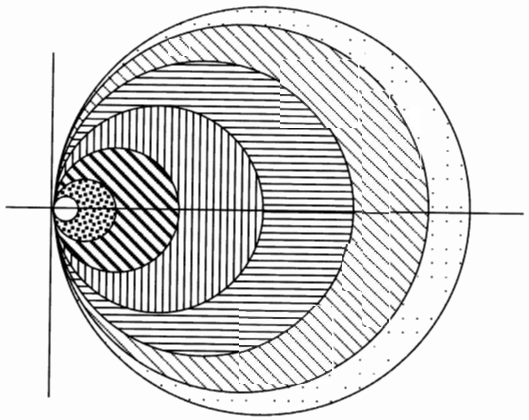


Figure 37 This nested sequence has empty intersection.

Proof Suppose that (b_n) is a sequence in fA. For each $n \in \mathbb{N}$ choose a point $a_n \in A$ such that $f(a_n) = b_n$. By compactness of A there exists a subsequence (a_{n_k}) that converges to some point $p \in A$. By continuity of f it follows that

$$b_{n_k} = f(a_{n_k}) \to fp \in fA.$$

Thus, any sequence (b_n) in fA has a subsequence converging to a limit in fA, and fA is compact. □

From Theorem 39 follows the natural generalization of the min/max theorem in Chapter 1 which concerned continuous real-valued functions defined on an interval $[a, b]$. See Theorem 23 in Chapter 1.

40 Corollary *A continuous real-valued function defined on a compact set is bounded; it assumes maximum and minimum values.*

Proof Let $f : M \to \mathbb{R}$ be continuous and let A be a compact subset of M. Theorem 39 implies that fA is a compact subset of $\mathbb{R}$, so by Theorem 29 it is closed and bounded. Thus, the greatest lower bound, v, and least upper bound, V, of fA exist and belong to fA; there exist points $p, P \in A$ such that for all $a \in A$, $v = fp \leq fa \leq fP = V$. □

Homeomorphisms and Compactness

A homeomorphism is a bi-continuous bijection. Originally, compactness was called bi-compactness. This is reflected in the next theorem.

41 Theorem *If M is compact and M is homeomorphic to N then N is compact. Compactness is a topological property.*

Proof If $f : M \to N$ is a homeomorphism then Theorem 39 implies that fM is compact. □

It follows that $[0, 1]$ and $\mathbb{R}$ are not homeomorphic. One is compact and the other is not.

42 Theorem *If M is compact then a continuous bijection $f : M \to N$ is a homeomorphism — its inverse bijection $f^{-1} : N \to M$ is continuous.*

Proof Suppose that $q_n \to q$ in N. Since f is a bijection $p_n = f^{-1}(q_n)$ and $p = f^{-1}(q)$ are well defined points in M. To check continuity of f^{-1} we must show that $p_n \to p$.

If (p_n) refuses to converge to p then there is a subsequence (p_{n_k}) and a $\delta > 0$ such that for all k, $d(p_{n_k}, p) \geq \delta$. Compactness of M gives a sub-subsequence $(p_{n_{k(\ell)}})$ that converges to a point $p^* \in M$ as $\ell \to \infty$.

Necessarily, $d(p, p^*) \geq \delta$, which implies that $p \neq p^*$. Since f is continuous

$$f(p_{n_{k(\ell)}}) \to f(p^*)$$

as $\ell \to \infty$. The limit of a convergent sequence is unchanged by passing to a subsequence, and so $f(p_{n_{k(\ell)}}) = q_{n_{k(\ell)}} \to q$ as $\ell \to \infty$. Thus, $f(p^*) = q = f(p)$, contrary to f being a bijection. It follows that $p_n \to p$ and therefore that f^{-1} is continuous. □

If M is not compact then Theorem 42 can fail. For example, the bijection $f : [0, 2\pi) \to \mathbb{R}^2$ defined by $f(x) = (\cos x, \sin x)$ is a continuous bijection onto the unit circle in the plane, but it is not a homeomorphism. This useful example was discussed on page 58. Not only does this f fail to be a homeomorphism, but there is no homeomorphism at all from $[0, 2\pi)$ to S^1. The circle is compact while $[0, 2\pi)$ is not. Therefore they are not homeomorphic.

Absolute Closedness

Not only is a compact space M closed in itself, as is every metric space, but it is also a closed subset of each metric space in which it is embedded. More precisely, we say that $h : M \to N$ **embeds** M into N if h is a homeomorphism from M onto hM. (The metric on hM is the one it inherits from N.) Topologically M and hM are equivalent. A property of M that holds

also for every embedded copy of M is an **absolute** or **intrinsic** property of M.

43 Theorem *A compact is absolutely closed and absolutely bounded.*

Proof Obvious from Theorems 29 and 39. □

For example, no matter how the circle is embedded in $\mathbb{R}^3$, it is a closed and bounded set. See also Exercises 28 and 87.

Uniform Continuity and Compactness

In Chapter 1 we defined the concept of uniform continuity for functions of a real variable. The definition in metric spaces is analogous. A function $f : M \to N$ is **uniformly continuous** if for each $\epsilon > 0$ there exists a $\delta > 0$ such that

$$p, q \in M \text{ and } d_M(p, q) < \delta \quad \Rightarrow \quad d_N(fp, fq) < \epsilon.$$

44 Theorem *Every continuous function defined on a compact is uniformly continuous.*

Proof Suppose not, and $f : M \to \mathbb{R}$ is continuous, M is compact, but f is not uniformly continuous. Then there is some $\epsilon > 0$ such that no matter how small δ is, there exist points $p, q \in M$ with $d(p, q) < \delta$ but $|fp - fq| \geq \epsilon$. Take $\delta = 1/n$ and let p_n, q_n be sequences of points in M such that $d(p_n, q_n) < 1/n$ while $|f(p_n) - f(q_n)| \geq \epsilon$. Compactness of M implies that there is a subsequence p_{n_k} which converges to some $p \in M$ as $k \to \infty$. Since $d(p_n, q_n) < 1/n \to 0$ as $n \to \infty$, (q_{n_k}) converges to the same limit as does (p_{n_k}) as $k \to \infty$; namely, $q_{n_k} \to p$. Continuity implies that $f(p_{n_k}) \to fp$ and $f(q_{n_k}) \to fp$. If k is large then

$$|f(p_{n_k}) - f(q_{n_k})| \leq |f(p_{n_k}) - fp| + |fp - f(q_{n_k})| < \epsilon,$$

contrary to the supposition that $|f(p_n) - f(q_n)| \geq \epsilon$ for all n. □

Theorem 44 gives a second proof that continuity implies uniform continuity on an interval $[a, b]$. For $[a, b]$ is compact.

3 Connectedness

As another application of these ideas, we consider the general notion of connectedness. Let A be a subset of M, a metric space. If A is neither the empty set nor is it M, then A is a **proper** subset of M. Recall that if A is both closed and open in M it is said to be clopen. The complement of a clopen set is clopen. The complement of a proper subset is proper.

If M has a proper clopen subset A, M is **disconnected**. For there is a **separation** of M into proper, disjoint clopen subsets,

$$M = A \sqcup A^c.$$

(The notation $\sqcup$ indicates disjoint union.) M is **connected** if it is not disconnected — it contains no proper clopen subset. Connectedness of M does not mean that M is connected *to* something, but rather that M is one unbroken thing. See Figure 38.

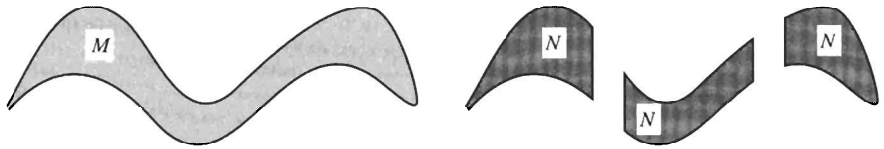


Figure 38 M and N illustrate the difference between being connected and being disconnected.

45 Theorem *If M is connected, $f : M \to N$ is continuous, and f is onto then N is connected. The continuous image of a connected is connected.*

Proof Simple! If A is a clopen proper subset of N then, according to the open and closed set conditions for continuity, $f^{\text{pre}}(A)$ is a clopen subset of M. Since f is onto and $A \neq \emptyset$, $f^{\text{pre}}(A) \neq \emptyset$. Similarly, $f^{\text{pre}}(A^c) \neq \emptyset$. Therefore $f^{\text{pre}}(A)$ is a proper clopen subset of M, contrary to M being connected. It follows that A can not exist and N is connected. □

46 Corollary *If M is connected and M is homeomorphic to N then N is connected. Connectedness is a topological property.*

Proof N is the continuous image of M. □

47 Corollary (Generalized Intermediate Value Theorem) *Every continuous real-valued function defined on a connected domain has the intermediate value property.*

Proof Assume that $f : M \to \mathbb{R}$ is continuous and M is connected. If f assumes values $\alpha < \beta$ in $\mathbb{R}$ and if it fails to assume some value γ with $\alpha < \gamma < \beta$, then

$$M = \{x \in M : f(x) < \gamma\} \sqcup \{x \in M : f(x) > \gamma\}$$

is a separation of M, contrary to connectedness. □

48 Theorem $\mathbb{R}$ *is connected.*

Proof If $U \subset \mathbb{R}$ is open, closed, and non-empty, we claim that $U = \mathbb{R}$. By Theorem 9, U is the countable disjoint union of open intervals whose endpoints do not belong to U. If (a, b) is one of these intervals and $b < \infty$, then closedness of U implies that $b \in U$, a contradiction, so $b = \infty$. Similarly, $a = -\infty$, and $U = \mathbb{R}$. Since $\mathbb{R}$ contains no proper clopen set, it is connected. □

49 Corollary (Intermediate Value Theorem for $\mathbb{R}$) *Every continuous function $f : \mathbb{R} \to \mathbb{R}$ has the intermediate value property.*

Proof Immediate from the Generalized Intermediate Value Theorem and connectedness of $\mathbb{R}$. □

Thus we have a second proof of the Intermediate Value Theorem 1.24.

50 Corollary *The following metric spaces are connected: the intervals (a, b), $[a, b]$, the circle, and all capital letters of the Roman alphabet.*

Proof The interval (a, b) is homeomorphic to $\mathbb{R}$, while $[a, b]$ is the continuous image of $\mathbb{R}$ under the map whose graph is shown in Figure 39. The circle is the continuous image of $\mathbb{R}$ under the map $t \mapsto (\cos t, \sin t)$. Connectedness of the letters A, ..., Z is equally clear. □

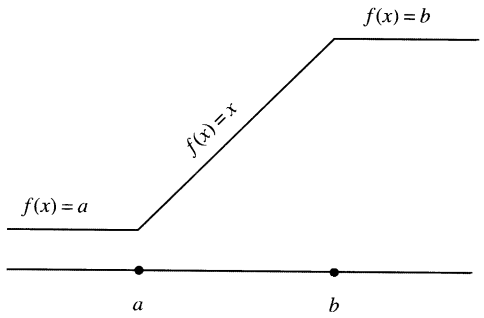


Figure 39 The function f surjects $\mathbb{R}$ continuously to $[a, b]$.

Connectedness properties give a good way to distinguish non-homeomorphic sets.

Example The union of two disjoint closed intervals is not homeomorphic to a single interval. One set is disconnected and the other is connected.

Example The closed interval $[a, b]$ is not homeomorphic to the circle S^1. For removal of a point $x \in (a, b)$ disconnects $[a, b]$ while the circle remains connected upon removal of any point. More precisely, suppose that $h : [a, b] \to S^1$ is a homeomorphism. Choose a point $x \in (a, b)$, and consider $X = [a, b] \setminus \{x\}$. The restriction of h to X is a homeomorphism from X onto Y, where Y is the circle with one point, hx, removed. But X is disconnected, while Y is connected. Hence h can not exist and the segment is not homeomorphic to the circle.

Example The circle is not homeomorphic to the figure eight. Removing any two points of the circle disconnects it, but this is not true of the figure eight. Or, removing the crossing point disconnects the figure eight, but removing no point disconnects the circle.

Example The circle is not homeomorphic to the disc. For removing two points disconnects the circle but does not disconnect the disc.

As you can see, it is useful to be able to recognize disconnected subsets S of a metric space M. By definition, S is a disconnected subset of M if it is disconnected when considered in its own right as a metric space with the metric it inherits from M; it has a separation $S = A \sqcup B$ such that A and B are proper closed subsets of S. According to the inheritance principle, closedness in S can be analyzed in M. The facts that no point of A is a limit of B and no point of b is a limit of A imply that the closure of B in M misses A and the closure of A in M misses B. (Note: "in M.") In symbols,

$$A \cap \overline{B} = \emptyset = \overline{A} \cap B.$$

The separated sets A, B must be disjoint but they need not be closed in M nor need their closures in M be disjoint.

Example The punctured interval $X = [a, b] \setminus \{c\}$ is disconnected if $a < c < b$. For $X = [a, c) \sqcup (c, b]$ is a separation of X. Observe that the closures in $\mathbb{R}$ of the separated sets have non-empty intersection.

Example Any subset Y of the punctured interval is disconnected if it meets both $[a, c)$ and $(c, b]$. For $Y = ([a, c) \cap Y) \sqcup ((c, b] \cap Y)$ is a separation of Y.

51 Theorem *The closure of a connected set is connected. More generally, if $S \subset T \subset \overline{S}$ and S is connected, then so is T.*

Proof It is equivalent to show that if T is disconnected then S is disconnected. Disconnectedness of T implies that

$$T = A \sqcup B$$

where A, B are clopen and proper. The set $K = A \cap S$ is clopen in S according to the Inheritance Theorem 13, but can K be improper? If $K = \emptyset$ then $A \subset S^c$. Since A is proper there exists $p \in A$. Since A is open in T, there exists a neighborhood $M_r p$ such that

$$T \cap M_r p \subset A \subset S^c.$$

The neighborhood $M_r p$ contains no points of S, which is contrary to p belonging to $\overline{S}$. Thus, $K \neq \emptyset$. Similarly, $L = B \cap S \neq \emptyset$, so $S = K \sqcup L$ is a separation of S, proving that S is disconnected. □

Example The outward spiral expressed in polar coordinates as

$$S = \{(r, \theta) : (1 - r)\theta = 1 \text{ and } \theta \geq \pi/2\}$$

has closure $\overline{S} = S \cup S^1$, S^1 being the unit circle. Since S is connected, so is the closure. See Figure 22.

52 Theorem *The union of connected sets sharing a common point is connected.*

Proof Let $S = \bigcup S_\alpha$ where each S_α is connected and $p \in \bigcap S_\alpha$. If S is disconnected then it has a separation $S = A \sqcup A^c$ where A, A^c are proper and clopen. One of them contains p, say it is A. Then $A \cap S_\alpha$ is a non-empty clopen subset of S_α. Since S_α is connected, $A \cap S_\alpha = S_\alpha$ for each α, and $A = S$. This implies that $A^c = \emptyset$, a contradiction. □

Example The 2-sphere S^2 is connected. For S^2 is the union of great circles, each passing through the poles.

Example Every convex set C in $\mathbb{R}^m$ (or in any metric space with a compatible linear structure) is connected. If we choose a point $p \in C$ then each $q \in C$ lies on a line segment $[p, q] \subset C$. Thus, C is the union of connected sets sharing the common point p. It is connected.

A **path** joining p to q in a metric space M is a continuous function $f : [a, b] \to M$ such that $fa = p$ and $fb = q$. If each pair of points in M can be joined by a path in M then M is **path-connected**. See Figure 40.

53 Theorem *Path-connected implies connected.*

Proof Assume that M is path-connected but not connected. Then $M = A \sqcup A^c$ for some proper clopen $A \subset M$. Choose $p \in A$ and $q \in A^c$. There is a path $f : [a, b] \to M$ from p to q. The sets $f^{\text{pre}}(A)$ and $f^{\text{pre}}(A^c)$ contradict connectedness of $[a, b]$. □

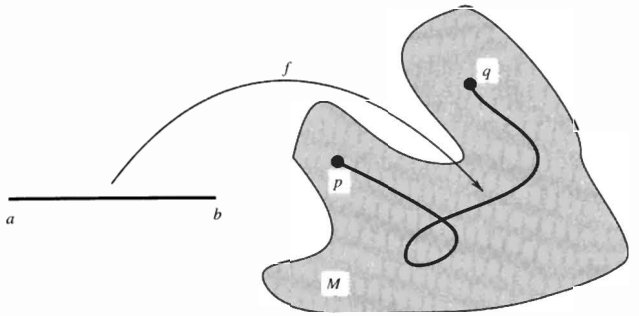


Figure 40 A path f in M that joins p to q.

Example All connected subsets of $\mathbb{R}$ are path-connected. See Exercise 65.

Example Every open connected subset of $\mathbb{R}^m$ is path-connected. See Exercises 58, 64.

Example The **topologist's sine curve** is a compact connected set that is not path-connected. It is $M = G \cup Y$ where

$$G = \{(x, y) \in \mathbb{R}^2 : y = \sin 1/x \text{ and } 0 < x \leq 1/\pi\}$$
$$Y = \{(0, y) \in \mathbb{R}^2 : -1 \leq y \leq 1\}.$$

See Figure 41. The metric on M is just Euclidean distance. Is M connected? Yes! The graph G is connected and M is its closure. By Theorem 51, M is connected.

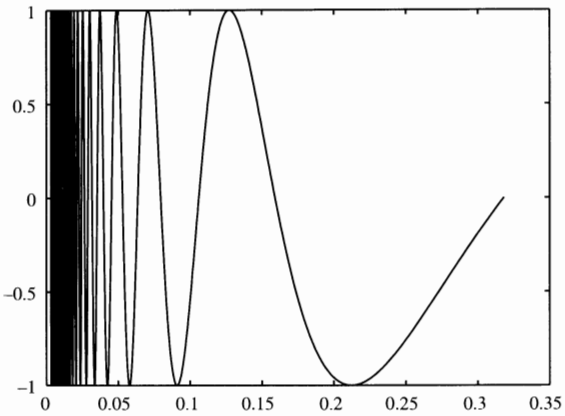


Figure 41 The topologist's sine curve M is a closed set. It includes the vertical segment Y at $x = 0$.

4 Coverings

For the sake of simplicity we have postponed discussing compactness in terms of open coverings until this point. Typically, students find coverings a challenging concept. It is central, however, to much of analysis, for example measure theory. A collection $\mathcal{U}$ of subsets of M covers $A \subset M$ if A is contained in the union of the sets belonging to $\mathcal{U}$. The collection $\mathcal{U}$ is a **covering** of A.

If $\mathcal{U}$ and $\mathcal{V}$ both cover A and if $\mathcal{V} \subset \mathcal{U}$ in the sense that each set $V \in \mathcal{V}$ belongs also to $\mathcal{U}$ then we say that $\mathcal{U}$ **reduces to** $\mathcal{V}$, and that $\mathcal{V}$ is a **subcovering**. If all the sets in $\mathcal{U}$ are open, $\mathcal{U}$ is an **open covering** of A. If every open covering of A reduces to a finite subcovering then we say that A is **covering compact**[†].

The idea is that if A is covering compact and $\mathcal{U}$ is an open covering of A then just a finite number of the open sets are actually doing the work of covering A. The rest are redundant.

A covering $\mathcal{U}$ of A is also called a cover of A. The members of $\mathcal{U}$ are *not* called covers. Instead, you could call them **scraps**. Imagine the covering as a patchwork quilt that covers a bed, the quilt being sewn together from overlapping scraps of cloth. See Figure 42.

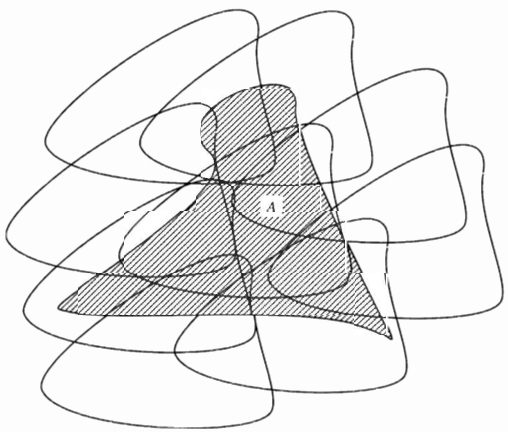


Figure 42 A covering of A by nine scraps.

The mere existence of a finite open covering of A is trivial and utterly worthless. *Every* set A has such a covering, namely the single open set M. Rather, for A to be covering compact, each and every open covering of A

[†] You will frequently find it said that an open covering of A *has* a finite subcovering. "Has" means "reduces to."

must reduce to a finite subcovering of A. Deciding directly whether this is so is daunting. How could you hope to verify the finite reducibility of all open coverings of A? There are so many of them. For this reason we concentrated on sequential compactness; it is relatively easy to check by inspection whether every sequence in a set has a convergent subsequence.

To check that a set is not covering compact it suffices to find an open covering which fails to reduce to a finite subcovering. Occasionally this is simple. For example, the set $(0, 1]$ is not covering compact in $\mathbb{R}$ because its covering

$$\mathcal{U} = \{(1/n, 1 + 1/n) : n \in \mathbb{N}\}$$

fails to reduce to a finite subcovering.

54 Theorem *For a subset A of a metric space M the following are equivalent:*

(a) A is covering compact.

(b) A is sequentially compact.

Proof that (a) implies (b) We assume that A is covering compact and prove that it is sequentially compact. Suppose not. There is a sequence (p_n) in A, no subsequence of which converges in A. Each point $a \in A$ therefore has some neighborhood $M_r a$ such that $p_n \in M_r a$ only finitely often. (The radius r may depend on the point a.) The collection $\{M_r a : a \in A\}$ is an open covering of A and by covering compactness it reduces to a finite subcovering

$$\{M_{r_1}(a_1), M_{r_2}(a_2), \ldots, M_{r_k}(a_k)\}$$

of A. Since p_n appears in each of these finitely many neighborhoods $M_{r_i}(a_i)$ only finitely often, it follows from the pigeon-hole principle that (p_n) has only finitely many terms, a contradiction. Thus, (p_n) can not exist and A is sequentially compact. □

The following presentation of the proof that (b) implies (a) appears in Royden's book, *Real Analysis*. A **Lebesgue number** for a covering $\mathcal{U}$ of A is a positive real number λ such that for each $a \in A$ there is some $U \in \mathcal{U}$ with $M_\lambda a \subset U$. Of course, the choice of this U depends on a. It is crucial, however, that the Lebesgue number λ is independent of $a \in A$.

The idea of a Lebesgue number is that we know each point $a \in A$ is contained in some $U \in \mathcal{U}$, and if λ is extremely small then $M_\lambda a$ is just a slightly swollen point — so the same should be true for it too. No matter where in A the neighborhood $M_\lambda a$ is placed, it lies wholly in some member of the covering. See Figure 43.

If A is noncompact then it may have open coverings with no positive Lebesgue number. For example, let A be the open interval $(0, 1)$. It is

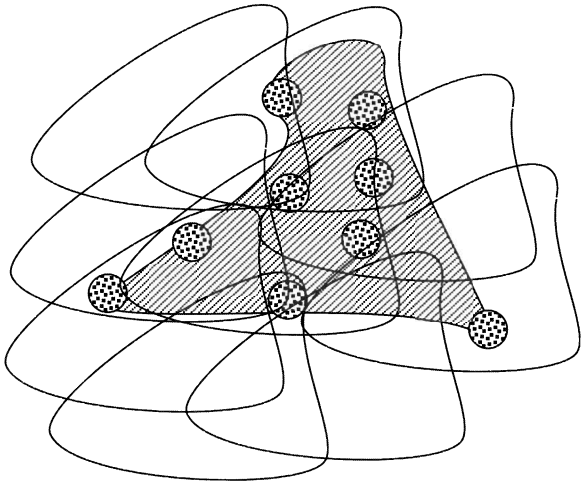


Figure 43 Small neighborhoods are like swollen points; $\mathcal{U}$ has a positive Lebesgue number.

covered by itself, $(0, 1) \subset (0, 1) = U$. Then for each $r > 0$, the r-neighborhood of a, $(a - r, a + r)$, fails to lie in U when $0 < a < r$. See Exercise 45.

55 Lebesgue Number Lemma *Every open covering of a sequentially compact set has a Lebesgue number* $\lambda > 0$.

Proof Suppose not: $\mathcal{U}$ is an open covering of a sequentially compact set A, and yet for each $\lambda > 0$ there exists an $a \in A$ such that no $U \in \mathcal{U}$ contains $M_\lambda a$. Take $\lambda = 1/n$ and let $a_n \in A$ be a point such that no $U \in \mathcal{U}$ contains $M_{1/n}(a_n)$. By sequential compactness, there is a subsequence (a_{n_k}) converging to some point $p \in A$. Since $\mathcal{U}$ is an open covering of A, there exist $r > 0$ and $U \in \mathcal{U}$ with $M_r p \subset U$. If k is large then $d(a_{n_k}, p) < r/2$ and $1/n_k < r/2$, which implies by the triangle inequality that

$$M_{1/n_k}(a_{n_k}) \subset M_r p \subset U,$$

contrary to the supposition that no $U \in \mathcal{U}$ contains $M_{1/n}(a_n)$. We conclude that, after all, $\mathcal{U}$ does have a Lebesgue number $\lambda > 0$. See Figure 44. □

Proof that (b) implies (a) in Theorem 54 Let $\mathcal{U}$ be an open covering of the sequentially compact set A. We want to reduce $\mathcal{U}$ to a finite subcovering. By the Lebesgue number Lemma, $\mathcal{U}$ has a Lebesgue number $\lambda > 0$. Choose any $a_1 \in A$ and some $U_1 \in \mathcal{U}$ such that

$$M_\lambda(a_1) \subset U_1.$$

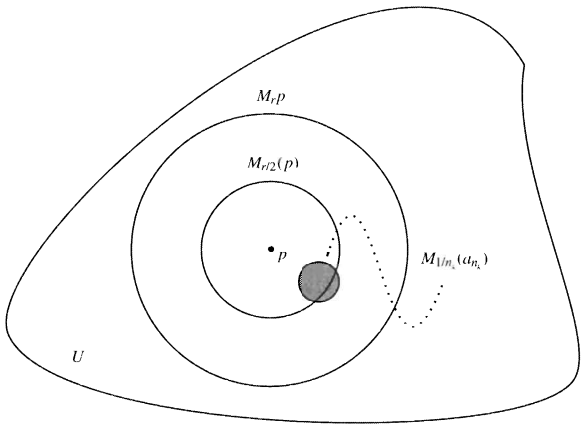


Figure 44 The neighborhood $M_r p$ engulfs the smaller neighborhood $M_{1/n_k}(a_{n_k})$.

If $U_1 \supset A$ then $\mathcal{U}$ reduces to the finite subcovering $\{U_1\}$ consisting of a single set, and the implication (b) $\Rightarrow$ (a) is proved. On the other hand, as is more likely, if U_1 does not contain A then we choose an uncovered point $a_2 \in A$ and a set $U_2 \in \mathcal{U}$ such that

$$M_\lambda(a_2) \subset U_2.$$

Either $\mathcal{U}$ reduces to the finite subcovering $\{U_1, U_2\}$ (and the proof is finished) or else we can continue, eventually producing a sequence (a_n) in A and a sequence (U_n) in $\mathcal{U}$ such that

$$M_\lambda(a_n) \subset U_n \text{ and } a_{n+1} \in (A \setminus (U_1 \cup \cdots \cup U_n)).$$

We will show that such sequences (a_n), (U_n) lead to a contradiction. By sequential compactness, there is a subsequence (a_{n_k}) that converges to some $p \in A$. For a large k, $d(a_{n_k}, p) < \lambda$ and

$$p \in M_\lambda(a_{n_k}) \subset U_{n_k}.$$

See Figure 45.

All a_{n_ℓ} with $\ell > k$ lie outside U_{n_k}, which contradicts their convergence to p. Thus, at some finite stage the process of choosing points a_n and sets U_n terminates, and $\mathcal{U}$ reduces to a finite subcovering $\{U_1, \ldots, U_n\}$ of A, which implies that A is covering compact. □

Upshot In light of Theorem 54, the term "compact" may now be applied equally to any set obeying (a) or (b).

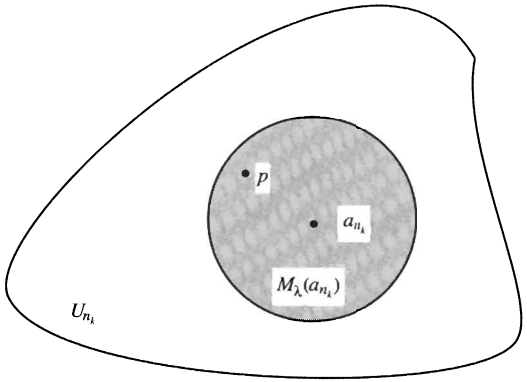


Figure 45 The point a_{n_k} is so near p that the neighborhood $M_\lambda(a_{n_k})$ engulfs p.

Total Boundedness

The Heine-Borel Theorem states that a subset of $\mathbb{R}^m$ is compact if and only if it is closed and bounded. In more general metric spaces, such as $\mathbb{Q}$, the assertion is false. But what if the metric space is complete? It is still false. For example the discrete metric on $\mathbb{N}$ makes $\mathbb{N}$ closed and bounded in $\mathbb{N}$ but it is noncompact: after all, what subsequence of $(1, 2, 3, \dots)$ converges?

But mathematicians do not quit easily. The Heine-Borel Theorem ought to generalize beyond $\mathbb{R}^m$ somehow. Here is the concept we need: a set $A \subset M$ is **totally bounded** if for each $\epsilon > 0$ there exists a finite covering of A by ϵ-neighborhoods. No mention is made of a covering reducing to a subcovering. How close total boundedness is to the worthless fact that every metric space has a finite open covering!

56 Theorem (Generalized Heine-Borel Theorem) *A subset of a complete metric space is compact if and only if it is closed and totally bounded.*

Proof Let A be a compact subset of M. Therefore it is closed. To see that it is totally bounded, let $\epsilon > 0$ be given and consider the ϵ-neighborhood covering of A,

$$\{M_\epsilon x : x \in A\}.$$

Compactness of A implies that this covering reduces to a finite subcovering and therefore A is totally bounded.

Conversely, assume that A is a closed and totally bounded subset of the complete metric space M. We claim that A is sequentially compact: any

sequence (a_n) in A has a subsequence that converges in A. Set $\epsilon_k = 1/k$, $k = 1, 2, \dots$. Since A is totally bounded we can cover it by finitely many ϵ_1-neighborhoods

$$M_{\epsilon_1}(q_1), \dots, M_{\epsilon_1}(q_m).$$

By the pigeon-hole principle, terms of the sequence a_n lie in at least one of these neighborhoods infinitely often, say it is $M_{\epsilon_1}(p_1)$. Choose

$$a_{n_1} \in A_1 = A \cap M_{\epsilon_1}(p_1).$$

Any subset of a totally bounded set is totally bounded, so we can cover A_1 by finitely many ϵ_2-neighborhoods. For one of them, say $M_{\epsilon_2}(p_2)$, a_n lies in $A_2 = A_1 \cap M_{\epsilon_2}(p_2)$ infinitely often. Choose $a_{n_2} \in A_2$ with $n_2 > n_1$.

Proceeding inductively, cover A_{k-1} by finitely many ϵ_k -neighborhoods, one of which, say $M_{\epsilon_k}(p_k)$, contains terms of the sequence (a_n) infinitely often. Then choose $a_{n_k} \in A_k = A_{k-1} \cap M_{\epsilon_k}(p_k)$ with $n_k > n_{k-1}$. Then (a_{n_k}) is a subsequence of (a_n). It is Cauchy because for $k, \ell \geq K$,

$$a_{n_k}, a_{n_\ell} \in A_K \quad \text{and} \quad \operatorname{diam} A_K \leq 2\epsilon_K = \frac{2}{K}.$$

Completeness of M implies that (a_{n_k}) converges to some $p \in M$ and since A is closed, $p \in A$. Hence A is compact. □

57 Corollary *A metric space is compact if and only if it is complete and totally bounded.*

Proof Every compact metric space M is complete. This is because, given a Cauchy sequence (p_n) in M, compactness implies that some subsequence converges in M, and if a Cauchy sequence has a convergent subsequence then the mother sequence converges too. As observed above, compactness immediately gives total boundedness.

Conversely, assume that M is complete and totally bounded. Any metric space is closed in itself. By Theorem 56, M is compact. □

Perfect Metric Spaces

A metric space M is **perfect** if $M' = M$: each $p \in M$ is a cluster point of M. Recall that M clusters at p if each $M_r p$ is an infinite set. For example $[a, b]$ is perfect and $\mathbb{Q}$ is perfect. $\mathbb{N}$ is not perfect; none of its points are cluster points.

58 Theorem *Every non-empty, perfect, complete metric space is uncountable.*

Proof Suppose not: assume M is non-empty, perfect, complete, and countable. Since M consists of cluster points it must be denumerable and not finite. Say

$$M = \{x_1, x_2, \dots\}$$

is a list of all the elements of M. Define

$$\widehat{M}_r p = \{q \in M : d(p, q) \leq r\}.$$

It is the "closed neighborhood" of radius r at p. If $x \notin Y$ we say that Y **excludes** x. Proceeding as in the proof of Theorem 56 we inductively choose a nested sequence of these closed neighborhoods to exclude more and more points in the sequence (x_n). They nest down to a point in M that is different from all the x_n, a contradiction.

Specifically take $y_1 = x_2$, $r_1 = \min(1, d(x_1, x_2)/2)$, and set

$$Y_1 = \widehat{M}_{r_1}(y_1).$$

Then Y_1 excludes x_1. Since M clusters at y_1, infinitely many x_n lie in $M_{r_1}(y_1)$, so we can choose $y_2 \in M_{r_1}(y_1)$ such that $y_2 \neq x_2$. The choice of

$$r_2 = \min\{1/2,\ d(y_2, x_2), (r_1 - d(y_1, y_2))\}$$

causes $Y_2 = \widehat{M}_{r_2}(y_2)$ to be a subset of Y_1 and to exclude x_1, x_2. See Figure 46.

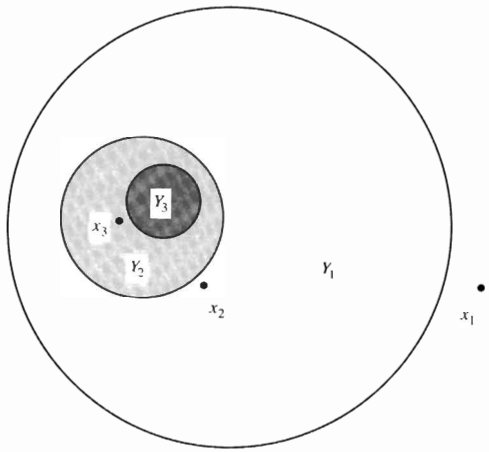


Figure 46 The exclusion of successively more points of the sequence (x_n).

Nothing stops us from continuing indefinitely, and we get a nested sequence of closed neighborhoods

$$Y_n = \widehat{M}_{r_n}(y_n),$$

such that Y_n excludes $x_1, \dots, x_n$, and has radius $r_n \leq 1/n$. Thus the center points y_n form a Cauchy sequence. Completeness of M implies that

$$\lim_{n\to\infty} y_n = y \in M$$

exists. Since the sets Y_n are closed and nested, $y \in Y_n$ for each n. Does y_n equal x_1? No, for Y_1 excludes x_1. Does it equal x_2? No, for Y_2 excludes x_2. In fact, for each n, $y \neq x_n$. The point y is nowhere in the supposedly complete list of elements of M, a contradiction. □

59 Corollary *$\mathbb{R}$ and $[a, b]$ are uncountable.*

Proof $\mathbb{R}$ is complete and perfect, while $[a, b]$ is compact, therefore complete, and perfect. Neither is empty. □

60 Corollary *A non-empty perfect complete metric space is everywhere uncountable in the sense that each r-neighborhood is uncountable.*

Proof The $r/2$-neighborhood $M_{r/2}(p)$ is perfect: it clusters at each of its points. The closure of a perfect set is perfect. Thus, $\overline{M_{r/2}(p)}$ is perfect. Being a closed subset of a complete metric space, it is complete. According to Theorem 58, $\overline{M_{r/2}(p)}$ is uncountable. Since $\overline{M_{r/2}(p)} \subset M_r p$, $M_r p$ is uncountable. □

5 Cantor Sets

Cantor sets are fascinating examples of compact sets that are maximally disconnected. (To emphasize the disconnectedness, one sometimes refers to a Cantor set as "Cantor dust.") Here is how to construct the standard Cantor set. Start with the unit interval $[0, 1]$ and remove its open middle third, $(1/3, 2/3)$. Then remove the open middle third from the remaining two intervals, and so on. This gives a nested sequence $C^0 \supset C^1 \supset C^2 \supset \dots$ where $C^0 = [0, 1]$, C^1 is the union of the two intervals $[0, 1/3]$ and $[2/3, 1]$, C^2 is the union of four intervals $[0, 1/9]$, $[2/9, 1/3]$, $[2/3, 7/9]$, and $[8/9, 1]$, C^3 is the union of eight intervals, and so on. See Figure 47.

In general C^n is the union of 2^n intervals, each of length $1/3^n$. Each C^n is compact. The **standard middle thirds Cantor set** is the nested intersection

$$C = \bigcap C^n.$$

We refer to C as "the" Cantor set. Clearly it contains the endpoints of each of the intervals comprising C^n. Actually, it contains uncountably many more points than these endpoints! There are other Cantor sets defined by

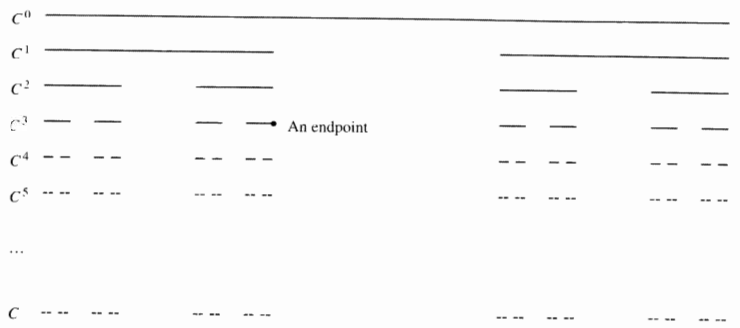


Figure 47 The construction of the standard middle thirds Cantor set C.

removing, say, middle fourths, pairs of middle tenths, etc. All Cantor sets turn out to be homeomorphic to the standard Cantor set. See Section 6.

A metric space M is **totally disconnected** if each point $p \in M$ has arbitrarily small clopen neighborhoods. That is, given $\epsilon > 0$ and $p \in M$, there exists a clopen set U such that

$$p \in U \subset M_\epsilon p.$$

For example, any discrete space is totally disconnected. So is $\mathbb{Q}$.

61 Theorem *The Cantor set is a compact, non-empty, perfect, and totally disconnected metric space.*

Proof The metric on C is the one it inherits from $\mathbb{R}$, the usual distance $|x - y|$. Because $C \supset E$, the set of endpoints of the C^n-intervals, C is non-empty and infinite. It is compact because it is the intersection of compacts.

To show C is perfect and totally disconnected, take any $x \in C$ and any $\epsilon > 0$. Fix n so large that $1/3^n < \epsilon$. The point x lies in one of the 2^n intervals I of length $1/3^n$ comprising C^n. The set of C-endpoints in I, $E \cap I$, is infinite and contained in the interval $(x-\epsilon, x+\epsilon)$. Thus C clusters at x and C is perfect. See Figure 48.

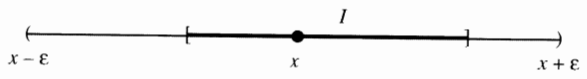


Figure 48 The endpoints of C cluster at x.

The interval I is closed in $\mathbb{R}$ and therefore in C^n. The complement $J = C^n \setminus I$ consists of finitely many closed intervals and is therefore closed too. Thus

$$C = (C \cap I) \sqcup (C \cap J)$$

exhibits C as the disjoint union of two clopen sets, $C \cap I$ being a subset of $(x - \epsilon, x + \epsilon)$ which contains x. Hence C is totally disconnected. □

62 Corollary *The Cantor set is uncountable.*

Proof Being compact, C is complete, and by Theorem 58, every complete, perfect, non-empty metric space is uncountable. □

A more direct way to see that the Cantor set is uncountable involves a geometric coding scheme. Take the code: 0 = left and 2 = right. Then

$$C_0 = \text{ left interval } = [0, 1/3] \quad C_2 = \text{ right interval } = [2/3, 1],$$

and $C^1 = C_0 \cup C_2$. Similarly, the left and right subintervals of C_0 are coded C_{00} and C_{02}, while the left and right subintervals of C_2 are C_{20} and C_{22}. This gives

$$C^2 = C_{00} \cup C_{02} \cup C_{20} \cup C_{22}.$$

The intervals that comprise C^3 are specified by strings of length 3. For instance, C_{220} is the left subinterval of C_{22}. In general an interval of C^n is coded by an **address string** of n symbols, each a 0 or a 2. Read it like a zip code. The first symbol gives the interval's gross location (left or right), the second symbol refines the location, the third refines it more, and so on.

Imagine now an **infinite address string** $\omega = \omega_1\omega_2\omega_3 \ldots$ of 0's and 2's. Corresponding to ω, we form a nested sequence of intervals

$$C_{\omega_1} \supset C_{\omega_1\omega_2} \supset C_{\omega_1\omega_2\omega_3} \supset \cdots \supset C_{\omega_1 \ldots \omega_n} \supset \ldots,$$

the intersection of which is a point $p = p(\omega) \in C$. Specifically,

$$p(\omega) = \bigcap_{n \in \mathbb{N}} C_{\omega|n}$$

where $\omega|n = \omega_1 \ldots \omega_n$ **truncates** ω to an address of length n. See Theorem 37.

As we have observed, any infinite address string defines a point in the Cantor set. Conversely, any point $p \in C$ has an address $\omega = \omega(p)$: its first n symbols $\alpha = \omega|n$ are specified by the interval C_α of C^n in which p lies. A second point q has a different address, since there is some n for which p and q lie in distinct intervals C_α and C_β of C^n.

In sum, the Cantor set is in one-to-one correspondence with the collection of addresses. Each address ω defines a point $p(\omega) \in C$ and each point $p \in C$ has a unique address $\omega(p)$.

If each 2 is replaced by 1, an address ω becomes a base-two expansion of some $x \in [0, 1]$. There are uncountably many $x \in [0, 1]$, uncountably many base-two expansions, and uncountably many addresses ω, all of which re-establishes the uncountability of C. In the same vein, ω can be interpreted directly as the base-three expansion of p. See Exercise 16 in Chapter 1. In Section 6 we will make more use of this geometric coding.

If $S \subset M$ and $\overline{S} = M$ then S is **dense** in M. For example, $\mathbb{Q}$ is dense in $\mathbb{R}$. The set S is **somewhere dense** if there exists an open non-empty set $U \subset M$ such that $\overline{S \cap U} \supset U$. If S is not somewhere dense then it is **nowhere dense**.

63 Theorem *The Cantor set contains no interval.*

Proof Suppose not: C contains some interval (a, b). Choose n such that $1/3^n < b - a$. The Cantor set is contained in a set C^n consisting of finitely many closed intervals, all shorter than $b - a$. Hence C^n can not contain (a, b), and neither can C. □

64 Corollary *The Cantor set is nowhere dense in* $\mathbb{R}$.

Proof If C is dense in an open set U then $U \subset \overline{C}$. Since C is closed, $U \subset C$. Any non-empty open set $U \subset \mathbb{R}$ contains an interval (a, b), but C never contains an interval. Hence $U = \emptyset$ and C is nowhere dense. □

The existence of an uncountable nowhere dense set is astonishing. Even more is true: the Cantor set is a **zero set**: it has "outer measure zero." By this we mean that, given any $\epsilon > 0$, there is a countable covering of C by open intervals (a_k, b_k), and the **total length** of the covering is

$$\sum_{k=1}^{\infty} b_k - a_k < \epsilon.$$

(Outer measure is one of the central concepts of Lebesgue Theory. See Chapter 7.) After all, C is a subset of C^n, which consists of 2^n closed intervals, each of length $1/3^n$. If n is large enough then $2^n/3^n < \epsilon$. Enlarging each of these closed intervals to an open interval keeps the sum of the lengths $< \epsilon$, and it follows that C is a zero set.

If we discard subintervals of [0, 1] in a different way, we can make a **fat Cantor set** — one that has positive outer measure. Instead of discarding the middle thirds of intervals at the n^{th} stage in the construction, we discard only the middle $1/n!$ portion. The discards are grossly smaller than the remaining intervals. See Figure 49. The total amount discarded from [0, 1] is < 1, and the total amount remaining, the outer measure of the fat Cantor set, is positive. See Exercise 3.32.

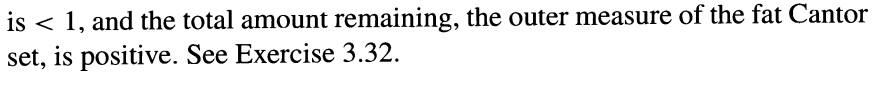


Figure 49 In forming a fat Cantor set, the gap intervals occupy a progressively smaller proportion of the Cantor set intervals.

6* Cantor Set Lore

In this section, we explore some arcane features of Cantor sets.

Although the continuous image of a connected set is connected, the continuous image of a disconnected set may well be connected. Just crush the disconnected set to a single point. Nevertheless, I hope you find the following result striking, for it means that the Cantor set C is the **universal compact metric space**, of which all others are merely shadows.

65 Cantor Surjection Theorem *Given a compact non-empty metric space M, there is a continuous surjection of C onto M.*

See Figure 50. Exercise 107 suggests a direct construction of a continuous surjection $C \to [0, 1]$, which is already an interesting fact. The proof of Theorem 65 involves a careful use of the address notation from Section 5 and the following simple lemma about dividing a metric space M into small pieces. A **piece** of M is any compact, non-empty subset of M.

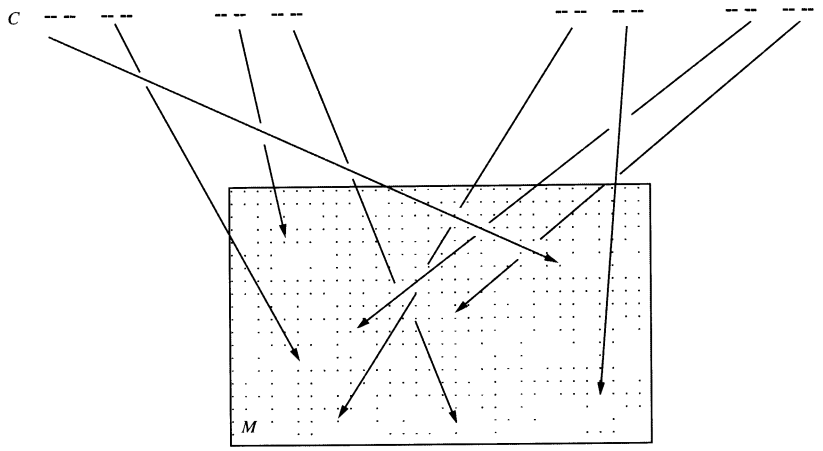


Figure 50 σ surjects C onto M.

66 Piece Lemma *A compact metric space M is the union of dyadically many small pieces. Specifically, given $\epsilon > 0$, there exist 2^k pieces of M, each with diameter $\leq \epsilon$, whose union is M.*

Proof (Recall that "dyadic" refers to powers of 2.) Cover M with neighborhoods of radius $\epsilon/2$. By compactness, finitely many of them suffice, say $U_1, \ldots, U_m$ cover M. Then M is the union of the m pieces

$$\overline{U_1}, \ldots, \overline{U_m},$$

each of diameter $\leq \epsilon$. Choose n with $m \leq 2^n$ and set $U_i = U_m$ for $m \leq i \leq 2^n$. (That is, repeat the last piece in the list $2^n - m$ times.) Then M is the union of the pieces $\overline{U_1}, \ldots, \overline{U_{2^n}}$, and each has diameter $\leq \epsilon$. □

Let us denote the length of an address string $\alpha = \alpha_1 \ldots \alpha_n$ by

$$|\alpha| = n.$$

There are exactly 2^n such α. We refer also to α as a **word** using the **letters** 0, 2. A **dyadic filtration** of M is a collection $\mathcal{M} = \{M_\alpha\}$ of pieces of M such that

(a) α varies freely in the collection of all finite words formed with the letters 0, 2.
(b) For each $n \in \mathbb{N}$, $M = \bigcup_{|\alpha|=n} M_\alpha$.
(c) If α is expressed as a **compound word** $\alpha = \beta\,\delta$ then $M_\alpha \subset M_\beta$.
(d) $\max\{\text{diam } M_\alpha : |\alpha| = n\} \to 0$ as $n \to \infty$.

We call $\mathcal{M}$ a filtration because its n^{th} level $\mathcal{M}_n = \{M_\alpha : |\alpha| = n\}$ consists of many small sets. Think of the sets M_α as the holes in a sieve that filters a liquid. (The analogy is imperfect since the M_α may overlap.) As $n \to \infty$, we filter M more and more finely.

67 Dyadic Filtration Lemma *Every non-empty compact metric space has a dyadic filtration.*

Proof By Lemma 66 there is an integer n_1 and 2^{n_1} pieces, each with diameter ≤ 1, whose union is M. Since there are 2^{n_1} dyadic words of length n_1, we use them to label these pieces as M_α. Then

$$M = \bigcup_{|\alpha|=n_1} M_\alpha.$$

If γ is a word of length $k < n_1$, define M_γ to be the union of all the pieces M_α such that $\gamma = \alpha|k$. (Recall that $\alpha|k$ is the truncation of α to its first

k letters.) This defines the filtration $\mathcal{M}_n$ for $1 \leq n \leq n_1$ and makes the nesting condition (c) automatic.

In the same way, each M_α can be expressed as a union of 2^{n_2} sub-pieces

$$M_\alpha = \bigcup_{|\beta|=n_2} (M_\alpha)_\beta,$$

each of diameter $\leq 1/2$. Define $M_{\alpha\beta} = (M_\alpha)_\beta$ when $|\beta| = n_2$. If γ is a word of length $k < n_2$ define $M_{\alpha\gamma}$ to be the union of all the $M_{\alpha\beta}$ with $\gamma = \beta|k$. See Figure 51.

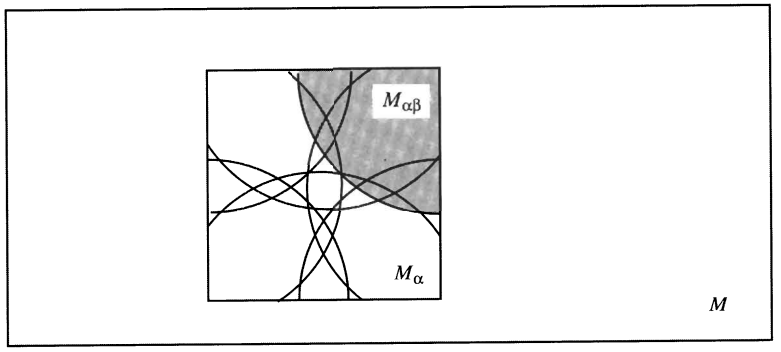


Figure 51 The piece M_α divided into sub-pieces $M_{\alpha\beta}$.

This defines pieces corresponding to all words δ of length $\leq n_1 + n_2$, and

$$M = \bigcup_{|\alpha|=n_1} \left(\bigcup_{|\beta|=n_2} (M_\alpha)_\beta \right) = \bigcup_{|\delta|=n_1+n_2} M_\delta.$$

For each word δ of length $n_1 + n_2$ is expressed uniquely as a compound word

$$\delta = \alpha\,\beta$$

where $|\alpha| = n_1$ and $|\beta| = n_2$. This extends the previous filtration to $\mathcal{M}_n$, $1 \leq n \leq n_1 + n_2$, and again (c) is automatic. Nothing stops us from passing to pieces of ever smaller diameter, which produces the desired dyadic filtration $\mathcal{M} = \bigcup \mathcal{M}_n$. □

Consider the intervals C_α that define the standard Cantor set C. Their intersections with C, $C_\alpha \cap C$, give a dyadic filtration of C. An infinite address $\omega = \omega_1\omega_2 \ldots$ specifies the precise location of a point $p \in C$ as

$$p(\omega) = \bigcap_{n\in\mathbb{N}} C_{\omega|n}.$$

As we saw in Section 5, points of C are in one-to-one correspondence with these infinite words ω. For each $p \in C$ there is an ω with $p = p(\omega)$, and for each ω, the point $p(\omega)$ belongs to C.

Proof of Theorem 65 We are given a compact, non-empty metric space M, and we must find a continuous surjection $\sigma : C \to M$, C being the standard Cantor set. Let $\{M_\alpha\}$ be a dyadic filtration of M and define

$$q(\omega) = \bigcap_{n \in \mathbb{N}} M_{\omega|n}$$

where ω is an infinite dyadic address. Then define $\sigma : C \to M$ by

$$\sigma(p) = q(\omega)$$

where $p = p(\omega)$.

Every point of M has at least one address. (It has several addresses if the pieces M_α overlap.) Hence σ is a surjection, and it remains to check continuity.

Let $\epsilon > 0$ be given and choose n so that

$$\max\{\text{diam } M_\alpha : |\alpha| = n\} < \epsilon.$$

Then choose δ so small that the intervals C_α in C^n lie farther apart than δ; i.e. $\delta < 1/3^n$. If $p, p' \in C$ and $|p - p'| < \delta$ then p, p' belong to a common interval C_α with $|\alpha| = n$, so their infinite addresses $\omega(p), \omega(p')$ both begin with the same string of length n, namely α. Therefore $\sigma(p)$, $\sigma(p')$ both belong to the same piece M_α, and

$$d(\sigma(p), \sigma(p')) < \epsilon$$

since M_α has diameter $< \epsilon$. This implies that σ is continuous. □

Peano Curves

68 Theorem *There exists a **Peano curve**, a continuous path in the plane which is **space filling** in the sense that its image has non-empty interior. In fact, there is a Peano curve whose image is the closed unit disc B^2.*

Proof Let $\sigma : C \to B^2$ be a continuous surjection supplied by Theorem 65. Extend σ to a map $\tau : [0, 1] \to B^2$ by setting

$$\tau(x) = \begin{cases} \sigma(x) & \text{if } x \in C \\ (1-t)\sigma(a) + t\sigma(b) & \text{if } x = (1-t)a + tb \in (a, b) \\ & \text{and } (a, b) \text{ is a gap interval.} \end{cases}$$

A **gap interval** is an interval $(a, b) \subset C^c$ such that $a, b \in C$. Because σ is continuous, $|\sigma(a) - \sigma(b)| \to 0$ as $|a - b| \to 0$. Hence τ is continuous. Its image includes the disc B^2 and thus has non-empty interior. In fact the image of τ is exactly B^2, since the disc is convex and τ just extends σ via linear interpolation. See Figure 52. □

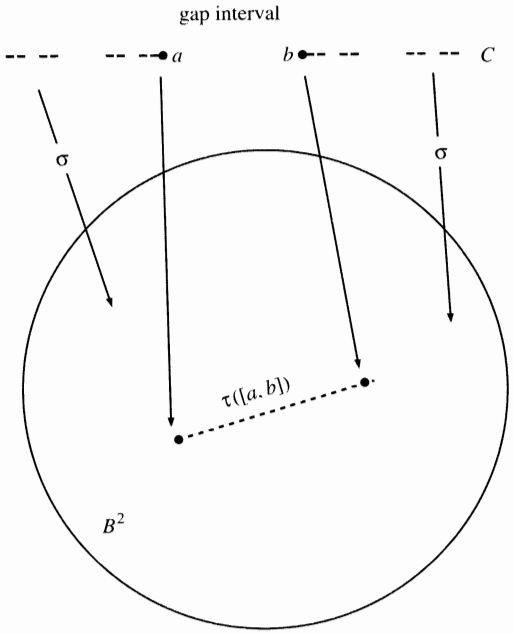


Figure 52 Filling in the Cantor surjection σ to make a Peano space filling curve τ.

This Peano curve can not be one-to-one since C is not homeomorphic to B^2. (C is disconnected while B^2 is connected.) In fact no Peano curve τ can be one-to-one since the removal of a point from $[0, 1]$ usually disconnects it, while this not true for open subsets of $\mathbb{R}^2$.

Cantor Spaces

We say that M is a **Cantor space** if, like the standard Cantor set C, it is compact, non-empty, perfect, and totally disconnected.

69 Moore-Kline Theorem *Every Cantor space is homeomorphic to the standard middle-thirds Cantor set C.*

A **Cantor piece** is a non-empty clopen subset S of a Cantor space M. It is easy to see that S is also a Cantor space. Since a Cantor space is totally

disconnected, each point has a small clopen neighborhood N. Thus, a Cantor space can always be divided into two disjoint Cantor pieces, $M = N \sqcup N^c$.

70 Dyadic Partition Lemma *A Cantor space is the disjoint dyadic union of small Cantor pieces.*

Proof A **partition** of a set is a division of it into disjoint subsets. In this case the small Cantor pieces form a partition of the Cantor space M.

Let $\epsilon > 0$ be given. Since M is totally disconnected and compact, we can cover it with finitely many clopen neighborhoods $U_1, \ldots, U_m$ having diameter $\leq \epsilon$. To make the sets U_i disjoint, define

$$\begin{aligned} V_1 &= U_1 \\ V_2 &= U_1 \setminus U_2 \\ &\cdots \\ V_m &= U_m \setminus (U_1 \cup \cdots \cup U_{m-1}). \end{aligned}$$

If any of these V_i is empty, discard it. The V_i are clopen and therefore are Cantor pieces.

Choose n with $m < 2^n$. Divide V_m into two disjoint Cantor pieces, W_1, X_1, divide X_1 into two disjoint Cantor pieces W_2, X_2, etc. Eventually this gives 2^n disjoint Cantor pieces

$$V_1, \ldots, V_{m-1}, W_1, \ldots, W_k, X_k,$$

whose union is M and whose diameters are $\leq \epsilon$. □

71 Dyadic Cantor Filtration Lemma *A Cantor space M has a dyadic filtration $\{M_\alpha\}$ by Cantor pieces. For each fixed $n \in \mathbb{N}$, the pieces M_α with $|\alpha| = n$ are disjoint.*

Proof Using Lemma 70 in place of Lemma 66, the proof of Lemma 67 gives a partition

$$M = \bigsqcup_{|\alpha|=n} M_\alpha,$$

where the symbol $\bigsqcup$ indicates disjoint union. □

Proof of the Moore-Kline Theorem 69 If we use the filtration supplied by Lemma 71, the continuous surjection $\sigma : C \to M$ defined in the proof of Theorem 65 is one-to-one. For disjointness of the Cantor pieces M_α with $|\alpha| = n$ implies that distinct points of M have distinct addresses. A continuous one-to-one surjection from one compact metric space to another is a homeomorphism. □

72 Corollary *The fat Cantor set is homeomorphic to the standard Cantor set.*

Proof Immediate from the Moore-Kline Theorem. □

73 Corollary *A Cantor set is homeomorphic to its own Cartesian square, $C \cong C \times C$.*

Proof It is enough to check that $C \times C$ is a Cantor space. It is. See Exercise 109. □

The fact that a non-trivial space is homeomorphic to its own Cartesian square is disturbing, is it not?

Ambient Topological Equivalence

Although all Cantor spaces are homeomorphic to each other when considered as abstract metric spaces, they can present themselves in very different ways as subsets of Euclidean space. Two sets A, B in $\mathbb{R}^m$ are **ambiently homeomorphic** if there is a homeomorphism of $\mathbb{R}^m$ to itself that sends A onto B. For example, the sets

$$A = \{0\} \cup [1, 2] \cup \{3\} \quad \text{and} \quad B = \{0\} \cup \{1\} \cup [2, 3]$$

are homeomorphic when considered as metric spaces, but there is no ambient homeomorphism of $\mathbb{R}$ that carries A to B. Similarly, the trefoil knot is homeomorphic but not ambiently homeomorphic to a planar circle. See also Exercises 125, 126.

74 Theorem *Any two Cantor sets in $\mathbb{R}$ are ambiently homeomorphic.*

Let M be a Cantor space contained in $\mathbb{R}$. According to Theorem 69, M is homeomorphic to the standard Cantor set C. We want to find a homeomorphism of $\mathbb{R}$ to itself that carries C to M.

The **convex hull** of $S \subset \mathbb{R}^m$ is the smallest convex set H that contains S. When $m = 1$, H is the smallest interval that contains S.

75 Lemma *A Cantor space $M \subset \mathbb{R}$ can be divided into two Cantor pieces whose convex hulls are disjoint.*

Proof Obvious from one-dimensionality of $\mathbb{R}$: choose a point $x \in \mathbb{R} \setminus M$ such that some points of M lie to the left of x and others lie to its right. Then

$$M \quad = \quad M \cap (-\infty, x) \quad \sqcup \quad (x, \infty) \cap M$$

divides M into disjoint Cantor pieces whose convex hulls are disjoint closed intervals. □

Proof of Theorem 74 Let $M \subset \mathbb{R}$ be a Cantor space. We will find a homeomorphism $\tau : \mathbb{R} \to \mathbb{R}$ sending C to M. Lemma 75 leads to a dyadic filtration $\{M_\alpha\}$ by Cantor pieces M_α whose convex hulls are disjoint when $|\alpha| = n$. With respect to the left/right order of $\mathbb{R}$, label the sets M_α in the same way that the Cantor middle third intervals are labeled: M_0 and M_2 are the left and right pieces of M, M_{00}, M_{02} are the left and right pieces of M_0, and so on. Then $\sigma : C \to M$ automatically is monotone. Extend σ across the gap intervals affinely as was done in the proof of Theorem 68, and extend it to $\mathbb{R} \setminus [0, 1]$ in any affine increasing fashion such that $\tau(0) = \sigma(0)$ and $\tau(1) = \sigma(1)$. Then $\tau : \mathbb{R} \to \mathbb{R}$ extends σ to $\mathbb{R}$. Monotonicity of σ implies that τ is one-to-one, while continuity of σ implies that τ is continuous. $\tau : \mathbb{R} \to \mathbb{R}$ is a homeomorphism that carries C onto M. □

As an example, one may construct a Cantor set in $\mathbb{R}$ by removing from $[0, 1]$ its middle third, then removing from each of the remaining intervals two symmetrically placed subintervals; then removing from each of the remaining six intervals, four symmetrically placed subintervals; and so forth. In the limit we get a nonstandard Cantor set M. According to Theorem 74, there is a homeomorphism of $\mathbb{R}$ to itself sending the standard Cantor set C to M.

Another example is the fat Cantor set mentioned on page 98.

Theorem *Every two Cantor spaces in $\mathbb{R}^2$ are ambiently homeomorphic.*

The key step is to show that M has a dyadic disc partition. That is, M can be divided into a dyadic number of Cantor pieces, each piece contained in the interior of a small topological disc D_i, the D_i being mutually disjoint. (A topological disc is any homeomorph of the closed unit disc B^2. Smallness refers to diam D_i.) The proofs I know of the existence of such dyadic partitions are tricky cut-and-paste arguments and are beyond the scope of this book. See Moise's book, *Geometric Topology in Dimensions 2 and 3*. See also Exercise 111.

Antoine's Necklace

A Cantor space $M \subset \mathbb{R}^m$ is **tame** if there is an ambient homeomorphism $h : \mathbb{R}^m \to \mathbb{R}^m$ that carries the standard Cantor set C (imagined to lie on the x_1-axis in $\mathbb{R}^m$) onto M. If M is not tame it is **wild**. Cantor spaces contained in the line or plane are tame. In 3-space, however, there are wild ones,

Cantor sets A so badly embedded in $\mathbb{R}^3$ that they act like curves. It is the lack of a "ball dyadic partition lemma" that causes the problem.

The first wild Cantor set was discovered by L. Antoine, and is known as **Antoine's Necklace**. The construction involves the solid torus or anchor ring, which is homeomorphic to the Cartesian product $B^2 \times S^1$. It is easy to imagine a necklace of solid tori: take an ordinary steel chain and modify it so its first and last links are also linked. See Figure 53.

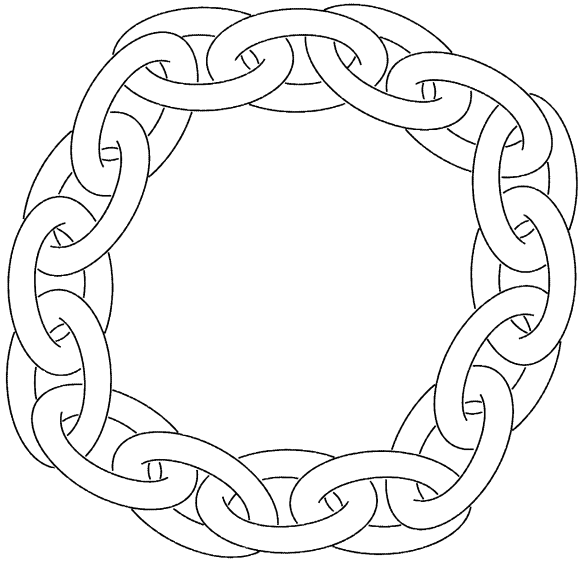


Figure 53 A necklace of sixteen solid tori.

Antoine's construction then goes like this. Draw a solid torus A^0. Interior to A^0, draw a necklace A^1 of several small solid tori, and make the necklace encircle the hole of A^0. Repeat the construction on each solid torus T comprising A^1. That is, interior to each T, draw a necklace of very small solid tori so that it encircles the hole of T. The result is a set $A^2 \subset A^1$ which is a necklace of necklaces. In Figure 53, A^2 would consist of 256 solid tori. Continue indefinitely, producing a nested decreasing sequence $A^0 \supset A^1 \supset A^2 \supset \ldots$. The set A^n is compact and consists of a large number (16^n) of extremely small solid tori arranged in a hierarchy of necklaces. It is an n^{th} order necklace. The intersection $A = \bigcap A^n$ is a Cantor space, since it is compact, perfect, non-empty, and totally disconnected. See Exercise 110.

Certainly A is bizarre, but is it wild? Is there no ambient homeomorphism h of $\mathbb{R}^3$ that sends the standard Cantor set C onto A? The reason that h can not exist is explained below.

Referring to Figure 54, the loop κ passing through the hole of A^0 can not be continuously shrunk to a point in $\mathbb{R}^3$ without hitting A. For if such a motion of κ avoids A then, by compactness, it also avoids one of the high order necklaces A^n. In $\mathbb{R}^3$ it is impossible to continuously de-link two linked loops, and it is also impossible to continuously de-link a loop from a necklace of loops. (These facts are intuitively believable but hard to prove. See Dale Rolfsen's book, *Knots and Links*.)

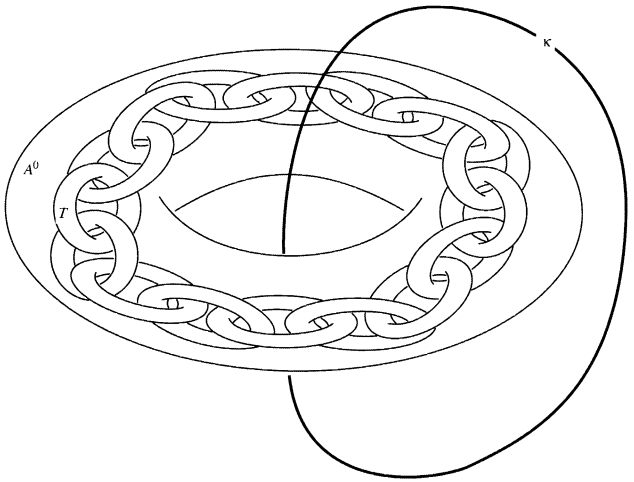


Figure 54 κ loops through A^0, which contains the necklace of solid tori.

On the other hand, each loop λ in $\mathbb{R}^3 \setminus C$ can be continuously shrunk to a point without hitting C. For there is no obstruction to pushing λ through the gap intervals of C.

Now suppose that there is an ambient homeomorphism h of $\mathbb{R}^3$ that sends C to A. Then $\lambda = h^{-1}(\kappa)$ is a loop in $\mathbb{R}^3 \setminus C$, and it can be shrunk to a point in $\mathbb{R}^3 \setminus C$, avoiding C. Applying h to this motion of λ continuously shrinks κ to a point, avoiding A, which we have indicated is impossible. Hence h can not exist, and A is wild.

7* Completion

Many metric spaces are complete (for example, all closed subsets of Euclidean space are complete), and completeness is a reasonable property to require of a metric space, especially in light of the following theorem.

76 Completion Theorem *Every metric space can be completed.*

This means that just as $\mathbb{R}$ completes $\mathbb{Q}$, we can take any metric space M and find a complete metric space $\widehat{M}$ containing M whose metric extends the metric of M. To put it another way, M is always a metric subspace of a complete metric space. In a natural sense the completion is uniquely determined by M.

77 Lemma *Given four points $p, q, x, y \in M$, we have*

$$|d(p,q) - d(x,y)| \leq d(p,x) + d(q,y).$$

Proof The triangle inequality implies that

$$d(x,y) \leq d(x,p) + d(p,q) + d(q,y)$$
$$d(p,q) \leq d(p,x) + d(x,y) + d(y,q),$$

and hence

$$-(d(p,x) + d(q,y)) \leq d(p,q) - d(x,y) \leq (d(p,x) + d(q,y)).$$

A number sandwiched between $-k$ and k has magnitude $\leq k$, which completes the proof. □

Proof of the Completion Theorem 76 We consider the collection $\mathscr{C}$ of all Cauchy sequences in M, convergent or not, and convert *it* into the completion of M. (This is a bold idea, is it not?) Cauchy sequences (p_n) and (q_n), are **co-Cauchy** if $d(p_n, q_n) \to 0$ as $n \to \infty$. Co-Cauchyness is an equivalence relation on $\mathscr{C}$. (This is easy to check.)

Define $\widehat{M}$ to be $\mathscr{C}$ modulo the equivalence relation of being co-Cauchy. Points of $\widehat{M}$ are equivalence classes $P = [(p_n)]$ such that (p_n) is a Cauchy sequence in M. The metric on $\widehat{M}$ is

$$D(P,Q) = \lim_{n\to\infty} d(p_n, q_n),$$

where $P = [(p_n)]$, $Q = [(q_n)]$. It only remains to verify three things:

(a) D is a well defined metric on $\widehat{M}$.
(b) $M \subset \widehat{M}$.
(c) $\widehat{M}$ is complete.

None of these assertions is really hard to prove, although the details are somewhat messy because of possible equivalence class/representative ambiguity.

(a) By Lemma 77

$$|d(p_m, q_m) - d(p_n, q_n)| \leq d(p_m, p_n) + d(q_m, q_n).$$

Thus $(d(p_n, q_n))$ is a Cauchy sequence in $\mathbb{R}$, and because $\mathbb{R}$ is complete,

$$L = \lim_{n\to\infty} d(p_n, q_n)$$

exists. Let (p'_n) and (q'_n) be sequences that are co-Cauchy with (p_n) and (q_n), and let

$$L' = \lim_{n\to\infty} d(p'_n, q'_n).$$

Then

$$\left|L - L'\right| \le |L - d(p_n, q_n)| + \left|d(p_n, q_n) - d(p'_n, q'_n)\right| + \left|d(p'_n, q'_n) - L'\right|.$$

As $n \to \infty$, the first and third terms tend to 0. By Lemma 77, the middle term is

$$\left|d(p_n, q_n) - d(p'_n, q'_n)\right| \le d(p_n, p'_n) + d(q_n, q'_n),$$

which also tends to 0 as $n \to \infty$. Hence $L = L'$ and D is well defined on $\widehat{M}$. The d distance on M is symmetric and satisfies the triangle inequality. Taking limits, these properties carry over to D on $\widehat{M}$, while positive definiteness follows directly from the co-Cauchy definition.

(b) Think of each $p \in M$ as a constant sequence, $\overline{p} = (p, p, p, p, \dots)$. Clearly it is Cauchy and clearly the D-distance between two constant sequences $\overline{p}$ and $\overline{q}$ is the same as the d-distance between the points p and q. In this way M is naturally a metric subspace of $\widehat{M}$.

(c) Let $(P_k)_{k\in\mathbb{N}}$ be a Cauchy sequence in $\widehat{M}$. We must find $Q \in \widehat{M}$ to which P_k converges as $k \to \infty$. (Note that (P_k) is a sequence of equivalence classes, not a sequence of points in M, and convergence refers to D not d.) Because D is well defined we can use a trick to shorten the proof. We claim that there exists an element $(p_{k,n}) \in P_k$ such that for all $m, n \in \mathbb{N}$,

$$d(p_{k,m}, p_{k,n}) < \frac{1}{k}. \tag{1}$$

All the terms of this representative are closely bunched, not merely those in the tail. Choose and fix some sequence $(p^*_{k,n})_{n\in\mathbb{N}}$ in P_k. Since it is Cauchy, there is some $N = N(k)$ such that for all $m, n \ge N$,

$$d(p^*_{k,m}, p^*_{k,n}) < \frac{1}{k}.$$

The sequence $p_{k,n} = p^*_{k,n+N}$ is co-Cauchy with $(p^*_{k,n})$ and it satisfies (1).

For each k, choose a representative $(p_{k,n}) \in P_k$ that satisfies (1), and define $q_n = p_{n,n}$. We claim that (q_n) is a Cauchy sequence in M and that P_k converges to its equivalence class Q as $k \to \infty$. That is, $\widehat{M}$ is complete.

Let $\epsilon > 0$ be given. There exists $N \ge 3/\epsilon$ such that if $k, \ell \ge N$

$$D(P_k, P_\ell) < \frac{\epsilon}{3}.$$

If $k, \ell \ge N$ then then by (1),

$$\begin{aligned} d(q_k, q_\ell) &= d(p_{k,k}, p_{\ell,\ell}) \\ &\le d(p_{k,k}, p_{k,n}) + d(p_{k,n}, p_{\ell,n}) + d(p_{\ell,n}, p_{\ell,\ell}) \\ &\le \frac{1}{k} + d(p_{k,n}, p_{\ell,n}) + \frac{1}{\ell} \\ &\le \frac{2\epsilon}{3} + d(p_{k,n}, p_{\ell,n}). \end{aligned}$$

The inequality is valid for all n and the left-hand side, $d(q_k, q_\ell)$, does not depend on n. The limit of $d(p_{k,n}, p_{\ell,n})$ as $n \to \infty$ is $D(P_k, P_\ell)$, which we know to be $< \epsilon/3$. Thus, if $k, \ell \ge N$ then $d(q_k, q_\ell) < \epsilon$ and (q_n) is Cauchy. Similarly we see that $P_k \to Q$ as $k \to \infty$. For, given $\epsilon > 0$, we choose $N \ge 2/\epsilon$ such that if $k, n \ge N$ then $d(q_k, q_n) < \epsilon/2$, from which it follows that

$$\begin{aligned} d(p_{k,n}, q_n) &\le d(p_{k,n}, p_{k,k}) + d(p_{k,k}, q_n) \\ &= d(p_{k,n}, p_{k,k}) + d(q_k, q_n) \\ &\le \frac{1}{k} + \frac{\epsilon}{2} < \epsilon. \end{aligned}$$

The limit of the left-hand side of this inequality, as $n \to \infty$, is $D(P_k, Q)$. Thus

$$\lim_{k\to\infty} P_k = Q$$

and $\widehat{M}$ is complete. □

Uniqueness of the completion is not surprising, and is left as Exercise 101. A different proof of the Completion Theorem is sketched in Exercise 4.37.

A Second Construction of $\mathbb{R}$ from $\mathbb{Q}$

In the particular case that the metric space M is $\mathbb{Q}$, the Completion Theorem leads to a construction of $\mathbb{R}$ from $\mathbb{Q}$ via Cauchy sequences. Note, however, that applying the theorem as it stands involves circular reasoning, for its proof uses completeness of $\mathbb{R}$ to define the metric D. Instead, we use only the Cauchy sequence strategy.

Convergence and Cauchyness for sequences of rational numbers are concepts that make perfect sense without a priori knowledge of $\mathbb{R}$. Just take all ϵ's and δ's in the definitions to be rational. The **Cauchy completion** $\widehat{\mathbb{Q}}$ of $\mathbb{Q}$ is the collection $\mathscr{C}$ of Cauchy sequences in $\mathbb{Q}$ modulo the equivalence relation of being co-Cauchy.

We claim that $\widehat{\mathbb{Q}}$ is a complete ordered field. The arithmetic on $\widehat{\mathbb{Q}}$ is defined by

$$P + Q = [(p_n + q_n)] \qquad P - Q = [(p_n - q_n)]$$
$$PQ = [(p_n q_n)] \qquad P/Q = [(p_n/q_n)]$$

where $P = [(p_n)]$ and $Q = [(q_n)]$. Of course $Q \neq \overline{0} = [(0, 0, \dots)]$ in the fraction P/Q. Exercise 102 asks you to check that these natural definitions make $\widehat{\mathbb{Q}}$ a field. Although there are many things to check — well definedness, commutativity, and so forth — all are effortless. There are no sixteen case proofs as with cuts. Also, just as with metric spaces, $\mathbb{Q}$ is naturally a subfield of $\widehat{\mathbb{Q}}$ when we think of $r \in \mathbb{Q}$ as the constant sequence $\overline{r} = [(r, r, \dots)]$.

That's the easy part; now the rest.

To define the order relation on $\widehat{\mathbb{Q}}$ we re-work some of the cut ideas. If $P \in \widehat{\mathbb{Q}}$ has a representative $[(p_n)]$, such that for some $\epsilon > 0$, we have $p_n \geq \epsilon$ for all n then P is positive. If $-P$ is positive, P is negative.

Then we define $P \prec Q$ if $Q - P$ is positive. Exercise 103 asks you to check that this defines an order on $\widehat{\mathbb{Q}}$, consistent with the standard order $<$ on $\mathbb{Q}$ in the sense that for all $p, q \in \mathbb{Q}$, $p < q \iff \overline{p} \prec \overline{q}$. In particular, you are asked to prove the trichotomy property: each $P \in \widehat{\mathbb{Q}}$ is either positive, negative, or zero, and these possibilities are mutually exclusive.

Combining Cauchyness with the definition of $\prec$ gives

$$\text{(2)} \qquad P = [(p_n)] \prec Q = [(q_n)] \iff \begin{array}{l} \text{there exist } \epsilon > 0 \text{ and } N \in \mathbb{N} \\ \text{such that for all } m, n \geq N, \\ p_m + \epsilon < q_n. \end{array}$$

It remains to check the least upper bound property. Let $\mathcal{P}$ be a non-empty subset of $\widehat{\mathbb{Q}}$ that is bounded above. We must find a least upper bound for $\mathcal{P}$. We assert that the least upper bound for $\mathcal{P}$ is the equivalence class Q of the following Cauchy sequence $(q_0, q_1, q_2, \dots)$.

(a) q_0 is the smallest integer such that $\overline{q_0}$ is an upper bound for $\mathcal{P}$.
(b) q_1 is the smallest fraction with denominator 2 such that $\overline{q_1}$ is an upper bound for $\mathcal{P}$.

(c) q_2 is the smallest fraction with denominator 4 such that $\overline{q_2}$ is an upper bound for $\mathcal{P}$.
(d) ...
(e) q_n is the smallest fraction with denominator 2^n such that $\overline{q_n}$ is an upper bound for $\mathcal{P}$.

Since $\mathcal{P} \neq \emptyset$ we can choose $P^* = [(p_n^*)] \in \mathcal{P}$. For some N^* and all $m, n \geq N^*$,

$$|p_m^* - p_n^*| < 1.$$

Hence $\overline{p_{N^*} - 1} \preceq P^*$. Integers smaller than $p_{N^*} - 1$ do not give upper bounds for $\mathcal{P}$, so q_0 is well defined. The other q_n are also well defined and form a monotone decreasing sequence

$$q_0 \geq q_1 \geq \cdots \geq q_n \geq \cdots$$

in $\mathbb{Q}$. By construction $|q_n - q_{n-1}| \leq 1/2^n$. Thus, if $m \leq n$ then

$$\begin{aligned} 0 \leq q_m - q_n &= q_m - q_{m+1} + q_{m+1} - q_{m+2} + \cdots + q_{n-1} - q_n \\ &\leq \frac{1}{2^{m+1}} + \cdots + \frac{1}{2^n} < \frac{1}{2^m}. \end{aligned}$$

It follows that (q_n) is Cauchy and $Q = [(q_n)] \in \widehat{\mathbb{Q}}$.

Suppose that Q is *not* an upper bound for $\mathcal{P}$. Then there is some $P = [(p_n)] \in \mathcal{P}$ with $Q \prec P$. By (2), there is an $\epsilon > 0$ and an N such that for all $n \geq N$,

$$q_N + \epsilon < p_n.$$

It follows that $\overline{q_N} \prec P$, a contradiction to $\overline{q_N}$ being an upper bound for $\mathcal{P}$.

On the other hand suppose that there is a smaller upper bound for $\mathcal{P}$, $R = (r_n) \prec Q$. By (2) there are $\epsilon > 0$ and N such that for all $m, n \geq N$,

$$r_m + \epsilon < q_n.$$

Fix a $k \geq N$ with $1/2^k < \epsilon$. Then for all $m \geq N$,

$$r_m < q_k - \epsilon < q_k - \frac{1}{2^k}.$$

By (2), $R \prec \overline{q_k - 1/2^k}$. Since R is an upper bound for $\mathcal{P}$, so is $\overline{q_k - 1/2^k}$, a contradiction to q_k being the *smallest* fraction with denominator 2^k such that $\overline{q_k}$ is an upper bound for $\mathcal{P}$.

This completes the verification that the Cauchy completion of $\mathbb{Q}$ is a complete ordered field. Uniqueness implies that it is isomorphic to the

complete ordered field $\mathbb{R}$ constructed by means of Dedekind cuts in Section 2 of Chapter 1. Decide for yourself which of the two constructions of the real number system you like better — cuts or Cauchy sequences. Cuts make least upper bounds straightforward and algebra awkward, while with Cauchy sequences, it is the reverse.

Exercises

1. $(0, 1)$ is an open subset of $\mathbb{R}$ but not of $\mathbb{R}^2$, when we think of $\mathbb{R}$ as the x-axis in $\mathbb{R}^2$. Prove this.
2. For which intervals $[a, b]$ in $\mathbb{R}$ is the intersection $[a, b] \cap \mathbb{Q}$ a clopen subset of the metric space $\mathbb{Q}$?
3. Prove directly from the definition of closed set that each single point is a closed subset of a metric space. Why does this imply that a finite set of points is also a closed set?
4. Prove that S clusters at p if and only if for each $r > 0$ there is a point $q \in M_r(p) \cap S$, such that $q \neq p$.
5. Prove that a set $U \subset M$ is open if and only if none of its points are limits of its complement.
6. If $S, T \subset M$, a metric space, and $S \subset T$, prove that
   (a) $\overline{S} \subset \overline{T}$.
   (b) $\text{int}(S) \subset \text{int}(T)$.
7. Construct a set with exactly three cluster points.
8. If $A \subset B \subset C$, A is dense in B, and B is dense in C prove that A is dense in C.
9. Is the set of dyadic rationals (the denominators are powers of 2) dense in $\mathbb{Q}$? In $\mathbb{R}$? Does one answer imply the other? (Recall that A is dense in B if $A \subset B$ and $\overline{A} \supset B$.)
10. (a) Find a metric space in which the boundary of $M_r p$ is not equal to the sphere of radius r at p, $\{x \in M : d(x, p) = r\}$.
    (b) Need the boundary be contained in the sphere?
11. Let $\mathcal{T}$ be the collection of open subsets of a metric space M, and $\mathcal{K}$ the collection of closed subsets. Show that there is a bijection from $\mathcal{T}$ onto $\mathcal{K}$.
12. Let M be a metric space with the discrete metric, or more generally a homeomorph of M.
    (a) Prove that every subset of M is clopen.
    (b) Prove that every function defined on M is continuous.
    (c) Which sequences converge in M?
13. Show that every subset of $\mathbb{N}$ is clopen. What does this tell you about every function $f : \mathbb{N} \to M$, where M is a metric space?
14. The **distance** from a point p in a metric space M to a non-empty subset $S \subset M$ is defined to be $\text{dist}(p, S) = \inf\{d(p, s) : s \in S\}$.
    (a) Show that p is a limit of S if and only if $\text{dist}(p, S) = 0$.
    (b) Show that $p \mapsto \text{dist}(p, S)$ is a uniformly continuous function of $p \in M$.

15. What is the set of points in $\mathbb{R}^3$ at distance exactly $1/2$ from the unit circle C in the plane,

$$T = \{p \in \mathbb{R}^3 : \exists q \in C \text{ and } d(p,q) = 1/2 \text{ and for all } q' \in C, d(p,q) \leq d(p,q')\}?$$

16. Show that $S \subset M$ is somewhere dense in M if and only if $\text{int}(\overline{S}) \neq \emptyset$. That is, S is nowhere dense in M if and only if its closure has empty interior.
17. Assume that $f : M \to N$ is a function from one metric space to another which satisfies the following condition: if a sequence (p_n) in M converges then the sequence $(f(p_n))$ in N converges. Prove that f is continuous.
18. The simplest type of mapping from one metric space to another is an **isometry**. It is a bijection $f : M \to N$ that preserves distance in the sense that for all $p, q \in M$,

$$d_N(fp, fq) = d_M(p, q).$$

If there exists an isometry from M to N then M and N are said to be isometric, $M \equiv N$. You might have two copies of a unit equilateral triangle, one centered at the origin and one centered elsewhere. They are isometric. Isometric metric spaces are indistinguishable as metric spaces.
   (a) Prove that every isometry is continuous.
   (b) Prove that every isometry is a homeomorphism.
   (c) Prove that [0, 1] is not isometric to [0, 2].
19. Prove that isometry is an equivalence relation: if M is isometric to N, show that N is isometric to M; show that any M is isometric to itself (what mapping of M to M is an isometry?); if M is isometric to N and N is isometric to P, show that M is isometric to P.
20. Is the perimeter of a square isometric to the circle? Homeomorphic? Explain.
21. Which capital letters of the roman alphabet are homeomorphic? Are any isometric? Explain.
22. Is $\mathbb{R}$ homeomorphic to $\mathbb{Q}$? Explain.
23. Is $\mathbb{Q}$ homeomorphic to $\mathbb{N}$? Explain.
24. An ant walks on the floor, ceiling, and walls of a cubical room. What metric is natural for the ant's view of its world? What metric would a spider consider natural? If the ant wants to walk from a point p to a point q, how could it determine the shortest path?

25. Assume that N is an open metric subspace of M and that $U \subset N$.
   (a) Prove that U is open in N if and only if it is open in M.
   (b) Conversely, prove that if openness of $S \subset N$ is equivalent to openness in M then N is open in M.
   (c) Do the same for closedness.
   (d) Deduce that a clopen metric subspace N is the only example in which the concepts of openness and closedness in the subspace agree exactly with the concepts in the big space.
26. Consider a sequence (x_n) in the metric space $\mathbb{R}$.
   (a) If (x_n) converges in $\mathbb{R}$ prove that the sequence of absolute values $(|x_n|)$ converges in $\mathbb{R}$.
   (b) Prove or disprove the converse.
27. Let (A_n) be a nested descreasing sequence of non-empty closed sets in the metric space M.
   (a) If M is complete and diam $A_n \to 0$ as $n \to \infty$, show that $\bigcap A_n$ is exactly one point.
   (b) To what assertions do the sets $[n, \infty)$ provide counterexamples?
28. Prove that there is an embedding of the line as a closed subset of the plane, and there is an embedding of the line as a bounded subset of the plane, but there is no embedding of the line as a closed and bounded subset of the plane.
29. (a) Prove that every convergent sequence is bounded. That is, if (p_n) converges in the metric space M, prove that there is some neighborhood $M_r q$ containing the set $\{p_n : n \in \mathbb{N}\}$.
   (b) Is the same true for a Cauchy sequence in an incomplete metric space?
30. A sequence (x_n) in $\mathbb{R}$ **increases** if $n < m$ implies $x_n \leq x_m$. It **strictly increases** if $n < m$ implies $x_n < x_m$. It **decreases** or **strictly decreases** if $n < m$ always implies $x_n \geq x_m$ or always implies $x_n > x_m$. A sequence is **monotone** if it increases or it decreases.
   (a) Prove that every sequence in $\mathbb{R}$ which is monotone and bounded converges in $\mathbb{R}$.
   (b) Prove that this monotone sequence condition is equivalent to the least upper bound property.
31. Let (x_n) be a sequence in $\mathbb{R}$.
   *(a) Prove that (x_n) has a monotone subsequence.
   (b) How can you deduce that any bounded sequence in $\mathbb{R}$ has a convergent subsequence?

32. Let (p_n) be a sequence and $f : \mathbb{N} \to \mathbb{N}$ a bijection. The sequence $(q_k)_{k\in\mathbb{N}}$ with $q_k = p_{f(k)}$ is a **rearrangement** of (p_n).
    (a) Are limits of a sequence unaffected by rearrangement?
    (b) What if f is an injection?
    (c) A surjection?
33. If $f : A \to B$ and $g : C \to B$ such that $A \subset C$ and for each $a \in A$, $f(a) = g(a)$ then f **extends to** g. Assume that $f : S \to \mathbb{R}$ is a uniformly continuous function defined on a subset S of a metric space M.
    (a) Prove that f extends to a uniformly continuous function $\overline{f} : \overline{S} \to \mathbb{R}$.
    (b) Prove that $\overline{f}$ is the unique continuous function defined on $\overline{S}$ such that $\overline{f}(x) = f(x)$ for all $x \in S$.
    (c) Prove the same things when N is a complete metric space and $f : S \to N$.
34. A map $f : M \to N$ is **open** if for each open set $U \subset M$, the image set $f(U)$ is open in N.
    (a) If f is open, is it continuous?
    (b) If f is a homeomorphism, is it open?
    (c) If f is an open, continuous bijection, is it a homeomorphism?
    (d) If $f : \mathbb{R} \to \mathbb{R}$ is a continuous surjection, must it be open?
    (e) If $f : \mathbb{R} \to \mathbb{R}$ is a continuous, open surjection, must it be a homeomorphism?
    (f) What happens in (e) if $\mathbb{R}$ is replaced by the unit circle S?
35. Fold a piece of paper in half.
    (a) Is this a continuous transformation of one rectangle into another?
    (b) Is it injective?
    (c) Draw an open set in the target rectangle, and find its pre-image in the original rectangle. Is it open?
    (d) What if the open set meets the crease?

    The **baker's transformation** is a similar mapping. A rectangle of dough is stretched to twice its length and then folded back on itself. Is the transformation continuous? A formula for the baker's transformation in one variable is $f(x) = 1 - |1 - 2x|$. The n^{th} iterate of f is $f^{\circ n} = f \circ f \circ \cdots \circ f$, n times. The **orbit** of a point x is

$$\{x, f(x), f \circ f(x), \ldots, f^{\circ n}(x), \ldots\}.$$

    (e) If x is rational prove that the orbit of x is a finite set.
    (f) If x is irrational what is the orbit?

36. Rotate the unit circle C by a fixed angle α, say $R : C \to C$. (In polar coordinates, the transformation R sends $(1, \theta)$ to $(1, \theta + \alpha)$.)
    (a) If α/π is rational, show that each orbit of R is a finite set.
    *(b) If α/π is irrational, show that each orbit is infinite and has closure equal to C.
37. Consider the identity map $id : C_{\max} \to C_{\text{int}}$ where $C_{\max}$ is the metric space $C([a, b], \mathbb{R})$ of continuous real valued functions defined on $[a, b]$, equipped with the max metric $d_{\max}(f, g) = \max |f(x) - g(x)|$, and C_{int} is $C([a, b], \mathbb{R})$ equipped with the integral metric,

$$d_{\text{int}}(f, g) = \int_a^b |f(x) - g(x)|\, dx.$$

    Show that id is a continuous linear bijection (an isomorphism) but its inverse is not continuous.
38. Let $\|\ \ \|$ be any norm on $\mathbb{R}^m$ and let $B = \{x \in \mathbb{R}^m : \|x\| \leq 1\}$. Prove that B is compact. [Hint: It suffices to show that B is closed and bounded with respect to the Euclidean metric.]
39. Assume that the Cartesian product of two non-empty sets $A \subset M$ and $B \subset N$ is compact in $M \times N$. Prove that A and B are compact.
40. Consider a function $f : M \to \mathbb{R}$. Its graph is the set

$$\{(p, y) \in M \times \mathbb{R} : y = fp\}.$$

    (a) Prove that if f is continuous then its graph is closed (as a subset of $M \times \mathbb{R}$).
    (b) Prove that if f is continuous and M is compact then its graph is compact.
    (c) Prove that if the graph of f is compact then f is continuous.
    (d) What if the graph is merely closed? Give an example of a discontinuous function $f : \mathbb{R} \to \mathbb{R}$ whose graph is closed.
41. Prove that the 2-sphere is not homeomorphic to the plane.
42. Draw a Cantor set C on the circle and consider the set A of all chords between points of C.
    (a) Prove that A is compact.
    ***(b) Is A convex?
43. Suppose that (K_n) is a nested sequence of compact non-empty sets, $K_1 \supset K_2 \supset \ldots$, and $K = \bigcap K_n$. If for some $\mu > 0$, each diam $K_n \geq \mu$, is it true that diam $K \geq \mu$?
44. Suppose that M is compact and that $\mathcal{U}$ is an open covering of M which is "redundant" in the sense that each $p \in M$ is contained in at

least two members of $\mathcal{U}$. Show that $\mathcal{U}$ reduces to a finite subcovering with the same property.

45. Suppose that every open covering of M has a positive Lebesgue number. Give an example of such an M that is not compact.

Exercises 46– 53 treat the basic theorems in the chapter, avoiding the use of sequences. The proofs will remain valid in general topological spaces.

46. Give a direct proof that $[a, b]$ is covering compact. [Hint: Let $\mathcal{U}$ be an open covering of $[a, b]$ and consider the set

$$C = \{x \in [a, b] : \text{ finitely many members of } \mathcal{U} \text{ cover } [a, x]\}.$$

Use the least upper bound principle to show that $b \in C$.]

47. Give a direct proof that a closed subset A of a covering compact set K is covering compact. [Hint: If $\mathcal{U}$ is an open covering of A, adjoin the set $W = M \setminus K$ to $\mathcal{U}$. Is $\mathcal{W} = \mathcal{U} \cup \{W\}$ an open covering of K? If so, so what?]
48. Give a proof of Theorem 39 using open coverings. That is, assume that A is a covering compact subset of M and $f : M \to N$ is continuous. Prove directly that fA is covering compact. [Hint: What is the criterion for continuity in terms of pre-images?]
49. Suppose that $f : M \to N$ is a continuous bijection and M is covering compact. Prove directly that f is a homeomorphism.
50. Suppose that M is covering compact and that $f : M \to N$ is continuous. Use the Lebesgue number lemma to prove that f is uniformly continuous. [Hint: Consider the covering of N by $\epsilon/2$-neighborhoods $\{N_{\epsilon/2}(q) : q \in N\}$ and its pre-image in M, $\{f^{\text{pre}}(N_{\epsilon/2}(q)) : q \in N\}$.]
51. Give a direct proof that the nested decreasing intersection of non-empty covering compact sets is non-empty. [Hint: If $A_1 \supset A_2 \supset \ldots$ are covering compact, consider the open sets $U_n = A_n^c$. If $\bigcap A_n = \emptyset$, what does $\{U_n\}$ cover?]
52. Generalize Exercise 51 as follows. Suppose that M is covering compact and $\mathscr{C}$ is a collection of closed subsets of M such that every intersection of finitely many members of $\mathscr{C}$ is non-empty. (Such a collection $\mathscr{C}$ is said to have the **finite intersection property**.) Prove that the **grand intersection** $\bigcap_{C \in \mathscr{C}} C$ is non-empty. [Hint: Consider the collection of open sets $\mathcal{U} = \{C^c : C \in \mathscr{C}\}$.]
53. If every collection of closed subsets of M which has the finite intersection property also has a non-empty grand intersection, prove that M is covering compact. [Hint: Given an open covering $\mathcal{U} = \{U_\alpha\}$, consider the collection of closed sets $\mathscr{C} = \{U_\alpha{}^c\}$.]

54. If S is connected, is the interior of S connected? Prove this or give a counter-example.
55. Theorem 51 states that the closure of a connected set is connected.
    (a) Is the closure of a disconnected set disconnected?
    (b) What about the interior?

*56. Prove that every countable metric space (not empty and not a singleton) is disconnected. [Astonishingly, there exists a countable topological space which is connected. Its topology does not arise from a metric.]

57. (a) Prove that a continuous function $f : M \to \mathbb{R}$, all of whose values are integers, is constant provided that M is connected.
    (b) What if all the values are irrational?
58. If U is an open subset of $\mathbb{R}^m$ and U is connected, prove that U is path-connected.
59. Prove that the annulus $A = \{z \in \mathbb{R}^2 : r \leq |z| \leq R\}$ is connected.
60. What function (given by a formula) is a homeomorphism from $(-1, 1)$ to $\mathbb{R}$? Is every open interval homeomorphic to $(0, 1)$? Why or why not?
61. A subset E of $\mathbb{R}^m$ is **starlike** if it contains a point p_0 (called a **center** for E) such that for each $q \in E$, the segment between p_0 and q lies in E.
    (a) If E is convex it is starlike.
    (b) Why is the converse false?
    (c) Is every starlike set connected?
    (d) Is every connected set starlike? Why or why not?

*62. Suppose that $E \subset \mathbb{R}^m$ is open, bounded, and starlike, and p_0 is a center for E.
    (a) Is it true or false that all points p_1 in a small enough neighborhood of p_0 are also centers for E?
    (b) Is the set of centers convex?
    (c) Is it closed?
    (d) Can it consist of a single point?
63. Suppose that $A, B \subset \mathbb{R}^2$ are convex, closed, and have non-empty interiors.
    (a) Prove that A, B are the closure of their interiors.
    (b) If A, B are compact, prove that they are homeomorphic.

[Hint: Draw a picture.]

64. (a) Prove that every connected open subset of $\mathbb{R}^m$ is path-connected.
    (b) Is the same true for open connected subsets of the circle?
    (c) What about connected non-open subsets of the circle?
65. List the convex subsets of $\mathbb{R}$ up to homeomorphism. How many are there and how many are compact?
66. List the closed convex sets in $\mathbb{R}^2$ up to homeomorphism. There are nine. How many are compact?
*67. Generalize Exercises 63 and 66 to $\mathbb{R}^3$; to $\mathbb{R}^m$.
68. Prove that (a, b) and $[a, b)$ are not homeomorphic metric spaces.
69. Let M and N be non-empty metric spaces.
    (a) If M and N are connected prove that $M \times N$ is connected.
    (b) What about the converse?
    (c) Answer the question again for path-connectedness.
70. Let H be the hyperbola $\{(x, y) \in \mathbb{R}^2 : xy = 1 \text{ and } x, y > 0\}$, and let X be the x-axis.
    (a) Is the set $S = X \cup H$ connected?
    (b) What if we replace H with the graph G of any continuous positive function $f : \mathbb{R} \to (0, \infty)$; is $X \cup G$ connected?
    (c) Give a counter-example if f is everywhere positive but discontinuous at just one point.
71. Assume that A, B are compact, disjoint, non-empty subsets of M. Prove that there exist points $a_0 \in A$, $b_0 \in B$ such that for all $a \in A$, and $b \in B$,

$$d(a_0, b_0) \leq d(a, b).$$

*72. An **arc** is a path with no self-intersection. Define the concept of arc connectedness and prove that a metric space is path-connected if and only if it is arc-connected.
73. The **Hawaiian earring** is the union of circles of radius $1/n$ and center $x = \pm 1/n$ on the x-axis, for $n \in \mathbb{N}$. See Figure 22 on page 53.
    (a) Is it connected?
    (b) Path-connected?
74. (a) The intersection of connected sets need not be connected. Give an example.
    (b) Suppose that $S_1, S_2, S_3, \ldots$ is a sequence of connected, closed subsets of the plane and $S_1 \supset S_2 \supset \ldots$. Is $S = \bigcap S_n$ connected? Give a proof or counter-example.
    *(c) Does the answer change if the sets are compact?
    (d) What is the situation for a nested decreasing sequence of compact path-connected sets?

75. Let $S = \mathbb{R}^2 \setminus \mathbb{Q}^2$. (Points $(x, y) \in S$ have at least one irrational coordinate.) Is S connected? Prove or disprove.
*76. The topologist's sine curve is the set

$$\{(x, y) : x = 0 \text{ and } |y| \leq 1, \text{ or } 0 < x \leq 1 \text{ and } y = \sin 1/x\}.$$

See Figure 41. The **topologist's sine circle** is shown in Figure 55. (It is the union of a circular arc and the topologist's sine curve.) Prove that it is path-connected, but not locally path-connected.

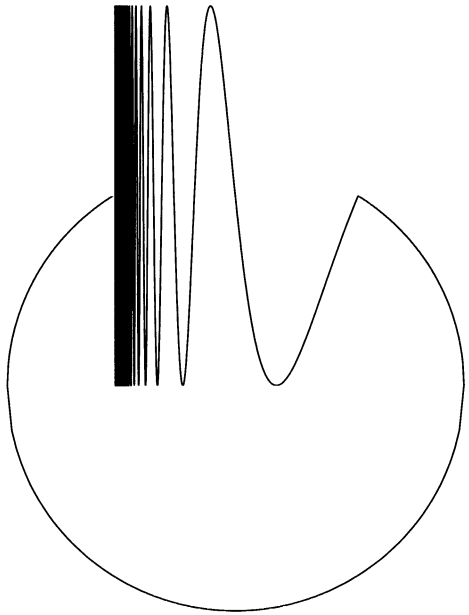


Figure 55 The topologist's sine circle.

77. If a metric space M is the union of path-connected sets S_α, all of which have the path-connected set K in common, is M path-connected?
78. $(p_1, \ldots, p_n)$ is an **ϵ-chain** in a metric space M if for each i, $p_i \in M$ and $d(p_i, p_{i+1}) < \epsilon$. The metric space is **chain-connected** if for each $\epsilon > 0$ and each pair of points $p, q \in M$ there is an ϵ-chain from p to q.
    (a) Show that every connected metric space is chain-connected.
    (b) Show that if M is compact and chain-connected then it is connected.
    (c) Is $\mathbb{R} \setminus \mathbb{Z}$ chain-connected?
    (d) If M is complete and chain-connected, is it connected?

79. Prove that if M is nonempty compact, locally path-connected and connected then it is path-connected.
80. The graph of $f : M \to \mathbb{R}$ is the set $\{(x, y) \in M \times \mathbb{R} : y = f(x)\}$. Since $M \times \mathbb{R}$ is a Cartesian product of two metric spaces it has a natural metric.
    (a) If M is connected and f is continuous, prove that the graph of f is connected.
    (b) Give an example to show that the converse is false.
    (c) If M is path-connected and f is continuous, show that the graph is path-connected.
    (d) What about the converse?
81. Consider the distance function $d : M \times M \to \mathbb{R}$.
    (a) Prove that d is continuous with respect to the natural sum metric defined on $M \times M$, namely $d_{\text{sum}}((p, q), (p', q')) = d(p, p') + d(q, q')$.
    (b) Is the same true for the other natural metrics d_E and $d_{\max}$?

    [You will find continuity of d useful in other exercises.]
82. Why is the sum metric on $\mathbb{R}^2$ called the **Manhattan metric** and the **taxicab metric**?
83. Let M_1, M_2 be metric spaces with metrics d_1, d_2, and let $M = M_1 \times M_2$ be their Cartesian product. Either
    (a) prove directly that the three natural metrics d_E, $d_{\max}$, and d_{sum} on M are metrics, or
    (b) prove in general that if $\|\ \ \|$ is any norm on $\mathbb{R}^2$ and we define a **distance vector** in $\mathbb{R}^2$ as $d_{pq} = (d_1(p_1, q_1), d_2(p_2, q_2))$ then $d(p, q) = \|d_{pq}\|$ is a metric on M. (The points $p = (p_1, p_2)$, $q = (q_1, q_2)$ belong to M.)
    (c) Does (b) imply (a)?

    [Hint: For (a) use Cauchy-Schwarz. For (b) use Euclidean geometry in $\mathbb{R}^2$ and the triangle inequality for the norm $\|\ \ \|$ to show that

$$d(p, r) = \|d_{pr}\| \leq \|d_{pq} + d_{qr}\| \leq \|d_{pq}\| + \|d_{qr}\| = d(p, q) + d(q, r)$$

    where $p, q, r \in M$.]
84. A metric space M with metric d can always be re-metrized so the metric becomes bounded. Simply define the **bounded metric**

$$\rho(p, q) = \frac{d(p, q)}{1 + d(p, q)}.$$

    (a) Prove that ρ is a metric. Why is it obviously bounded?

    (b) Prove that the identity map $M \to M$ is a homeomorphism from M with the d-metric to M with the ρ-metric.
    (c) Infer that boundedness of M is not a topological property.
    (d) Find homeomorphic metric spaces, one bounded and the other not.

*85. The implications of compactness are frequently equivalent to it. Prove
    (a) If every continuous function $f : M \to \mathbb{R}$ is bounded then M is compact.
    (b) If every continuous bounded function $f : M \to \mathbb{R}$ achieves a maximum or minimum, then M is compact.
    (c) If every continuous function $f : M \to \mathbb{R}$ has compact range fM, then M is compact.
    (d) If every nested decreasing sequence of non-empty closed subsets of M has non-empty intersection, then M is compact.

    Together with Theorems 54 and 56, (a) – (d) give seven equivalent definitions of compactness. [Hint: Reason contrapositively. If M is not compact it contains a sequence (p_n) that has no convergent subsequence. It is fair to assume that the points p_n are distinct. Find radii $r_n > 0$ such that the neighborhoods $M_{r_n}(p_n)$ are disjoint and no sequence $q_n \in M_{r_n}(p_n)$ has a convergent subsequence. Using the metric define a function $f_n : M_{r_n}(p_n) \to \mathbb{R}$ with a spike at p_n, such as

$$f_n(x) = \frac{r_n - d(x, p_n)}{a_n + d(x, p_n)}$$

    where $a_n > 0$. Set $f(x) = f_n(x)$ if $x \in M_{r_n}(p_n)$, and $f(x) = 0$ if x belongs to no $M_{r_n}(p_n)$. Show that f is continuous. With the right choice of a_n show that f is unbounded. With a different choice of a_n, it is bounded but achieves no maximum, and so on.]
86. Let M be a metric space on which the metric d is bounded by 1, $d(p, q) < 1$ for all $p, q \in M$. The **cone** over M is the set

$$C \quad = \quad C(M) \quad = \quad \{p_0\} \ \cup \ (0, 1] \times M$$

    with the polar metric

$$\rho((r, p), (s, q)) = |r - s| + \min\{r, s\} d(p, q)$$
$$\rho((r, p), p_0) = r.$$

    The point p_0 is the vertex of the cone. Prove that ρ is a metric on C.
87. Recall that if for each embedding of M, $h : M \to N$, hM is closed in N then M is said to be absolutely closed. If each hM is bounded

then M is absolutely bounded. Theorem 43 implies that compact sets are absolutely closed and absolutely bounded. Prove:

(a) If M is absolutely bounded, then M is compact.

*(b) If M is absolutely closed, then M is compact.

Thus, these are two more conditions equivalent to compactness. [Hint: From Exercise 85(a), if M is noncompact there is a continuous function $f : M \to \mathbb{R}$ that is unbounded. For Exercise 87(a), show that $F(x) = (x, f(x))$ embeds M onto a nonbounded subset of $M \times \mathbb{R}$. For 87(b), justify the additional assumption that the metric on M is bounded by 1. Then use Exercise 85(b) to show that if M is noncompact there is a continuous function $g : M \to (0, 1]$ such that for some nonclustering sequence (p_n), we have $g(p_n) \to 0$ as $n \to \infty$. Finally, show that $G(x) = (g(x), x)$ embeds M onto a nonclosed subset of the cone $C(M)$ discussed in Exercise 86.]

88. (a) Prove that every function defined on a discrete metric space is uniformly continuous.
    (b) Infer that it is false to assert that if every continuous function $f : M \to \mathbb{R}$ is uniformly continuous, then M is compact.
    (c) Prove, however, that if M is a subset of a compact metric space K and every continuous function $f : M \to \mathbb{R}$ is uniformly continuous, then M is compact.

89. Recall that p is a **cluster point** of S if each $M_r p$ contains infinitely many points of S. The set of cluster points of S is denoted as S'. Prove:
    (a) If $S \subset T$ then $S' \subset T'$.
    (b) $(S \cup T)' = S' \cup T'$.
    (c) $S' = (\overline{S})'$.
    (d) S' is closed in M; that is, $S'' \subset S'$ where $S'' = (S')'$.
    (e) Calculate $\mathbb{N}'$, $\mathbb{Q}'$, $\mathbb{R}'$, $(\mathbb{R} \setminus \mathbb{Q})'$, $\mathbb{Q}''$.
    (f) Let T be the set of points $\{1/n : n \in \mathbb{N}\}$. Calculate T' and T''.
    (g) Give an example showing that S'' can be a proper subset of S'.

90. Recall that p is a **condensation point** of S if each $M_r p$ contains uncountably many points of S. The set of condensation points of S is denoted as S^*. Prove:
    (a) If $S \subset T$ then $S^* \subset T^*$.
    (b) $(S \cup T)^* = S^* \cup T^*$.
    (c) $S^* \subset \overline{S}^*$ where $\overline{S}^* = (\overline{S})^*$
    (d) S^* is closed in M; that is, $S^{*\prime} \subset S^*$ where $S^{*\prime} = (S^*)'$.
    (e) $S^{**} \subset S^*$ where $S^{**} = (S^*)^*$.

(f) Calculate $\mathbb{N}^*$, $\mathbb{Q}^*$, $\mathbb{R}^*$, $(\mathbb{R} \setminus \mathbb{Q})^*$, $\mathbb{Q}^{**}$.

(g) Give an example showing that S^* can be a proper subset of $(\overline{S})^*$. Thus, (c) is not in general an equality.

**(h) Give an example that S^{**} can be a proper subset of S^*. Thus, (e) is not in general an equality. [Hint: Consider the set M of all functions $f : [a, b] \to [0, 1]$, continuous or not, and let the metric on M be the sup metric, $d(f, g) = \sup\{|f(x) - g(x)| : x \in [a, b]\}$. Consider the set S of all "δ-functions with rational values."]

**(i) Give examples that show in general that S^* neither contains nor is contained in S'^* where $S'^* = (S')^*$. [Hint: δ-functions with values $1/n$, $n \in \mathbb{N}$.]

91. Recall that p is an interior point of $S \subset M$ if some $M_r p$ is contained in S. The set of interior points of S is the **interior** of S and is denoted $\operatorname{int} S$. For all subsets S, T of the metric space M prove:
    (a) $\operatorname{int} S = S \setminus \partial S$.
    (b) $\operatorname{int} S = (\overline{S^c})^c$.
    (c) $\operatorname{int}(\operatorname{int} S) = \operatorname{int} S$.
    (d) $\operatorname{int}(S \cap T) = \operatorname{int} S \cap \operatorname{int} T$.
    (e) What are the dual equations for the closure?
    (f) Prove that $\operatorname{int}(S \cup T) \supset (\operatorname{int} S \cup \operatorname{int} T)$. Show by example that the inclusion can be strict, i.e., not an equality.

92. A point p is a boundary point of a set $S \subset M$ if every neighborhood $M_r p$ contains points of both S and S^c. The **boundary** of S is denoted ∂S. For all subsets S, T of a metric space M prove:
    (a) S is clopen if and only if $\partial S = \emptyset$.
    (b) $\partial S = \partial S^c$.
    (c) $\partial\partial S \subset \partial S$.
    (d) $\partial\partial\partial S = \partial\partial S$.
    (e) $\partial(S \cup T) \subset \partial S \cup \partial T$.
    (f) Give an example in which (c) is a strict inclusion, $\partial\partial S \neq \partial S$.
    (g) What about (e)?

*93. Suppose that E is an uncountable subset of $\mathbb{R}$. Prove that there exists a point $p \in \mathbb{R}$ at which E condenses. [Hint: Use decimal expansions. Why must there be an interval $[n, n+1)$ containing uncountably many points of E? Why must it contain a decimal subinterval with the same property? (A decimal subinterval $[a, b)$ has endpoints $a = n + k/10$, $b = n + (k+1)/10$ for some digit k, $0 \leq k \leq 9$.) Do you see lurking

the decimal expansion of a condensation point?] Generalize to $\mathbb{R}^2$ and to $\mathbb{R}^m$.

94. The metric space M is **separable** if it contains a countable dense subset. [Note the confusion of language: "Separable" has nothing to do with "separation."]
    (a) Prove that $\mathbb{R}^m$ is separable.
    (b) Prove that every compact metric space is separable.

95. *(a) Prove that any metric subspace of a separable metric space is separable, and deduce that any metric subspace of $\mathbb{R}^m$ or of a compact metric space is separable.
    (b) Is the property of being separable topological?
    (c) Is the continuous image of a separable metric space separable?

96. Think up a non-separable metric space.

97. Let $\mathcal{B}$ denote the collection of all ϵ-neighborhoods in $\mathbb{R}^m$ whose radius ϵ is rational and whose center has all coordinates rational.
    (a) Prove that $\mathcal{B}$ is countable.
    (b) Prove that every open subset of $\mathbb{R}^m$ can be expressed as the countable union of members of $\mathcal{B}$.

    (The union need not be disjoint, but it is at most a countable union because there are only countably many members of $\mathcal{B}$. A collection such as $\mathcal{B}$ is called a **countable base** for the topology of $\mathbb{R}^m$.)

98. (a) Prove that any separable metric space has a countable base for its topology, and conversely that any metric space with a countable base for its topology is separable.
    (b) Infer that every compact metric space has a countable base for its topology.

*99. Referring to Exercise 90, assume now that M is separable, $S \subset M$, and, as before S' is the set of cluster points of S while S^* is the set of condensation points of S. Prove:
    (a) $S^* \subset (S')^* = (\overline{S})^*$.
    (b) $S^{**} = S^{*\prime} = S^*$.
    (c) Why is (a) not in general an equality?

    [Hints: For (a) write $S \subset (S \setminus S') \cup S'$ and $\overline{S} = (S \setminus S') \cup S'$, show that $(S \setminus S')^* = \emptyset$, and use Exercise 90(a). For (b), Exercise 90(d) implies that $S^{**} \subset S^{*\prime} \subset S^*$. To prove that $S^* \subset S^{**}$, write $S \subset (S \setminus S^*) \cup S^*$ and show that $(S \setminus S^*)^* = \emptyset$.]

*100. Prove that
    (a) An uncountable subset of $\mathbb{R}$ clusters at some point of $\mathbb{R}$.
    (b) An uncountable subset of $\mathbb{R}$ clusters at some point of itself.

    (c) An uncountable subset of $\mathbb{R}$ condenses at uncountably many points of itself.
    (d) What about $\mathbb{R}^m$?
    (e) What about any compact metric space?
    (f) What about any separable metric space?

101. Prove that the completion of a metric space is unique in the following natural sense: A completion of a metric space M is a complete metric X space containing M as a metric subspace such that M is dense in X. That is, every point of X is a limit of M.
    (a) Prove that M is dense in the completion $\widehat{M}$ constructed in the proof of Theorem 76.
    (b) If X and X' are two completions of M prove that there is an isometry $i : X \to X'$ such that $i(p) = p$ for all $p \in M$.
    (c) Prove that i is the unique such isometry.
    (d) Infer that $\widehat{M}$ is unique.

*102. Prove that $\widehat{\mathbb{Q}}$, the Cauchy sequences in $\mathbb{Q}$ modulo the equivalence relation of being co-Cauchy, is a field with respect to the natural arithmetic operations defined on page 112, and that $\mathbb{Q}$ is naturally a subfield of $\widehat{\mathbb{Q}}$.

103. Prove that the order on $\widehat{\mathbb{Q}}$ defined on page 112 is a bona fide order which agrees with the standard order on $\mathbb{Q}$.

*104. Let Σ be the set of all infinite sequences of 0's and 1's. For example, $(100111000011111\ldots) \in \Sigma$. Define the metric

$$d(a, b) = \sum \frac{|a_n - b_n|}{2^n}$$

    where $a = (a_n)$ and $b = (b_n)$ are points in Σ.
    (a) Prove that Σ is compact.
    (b) Prove that Σ is homeomorphic to the Cantor set.

*105. A metric on M is an **ultrametric** if for all $x, y, z \in M$,

$$d(x, z) \leq \max\{d(x, y),\ d(y, z)\}.$$

    (Intuitively this means that the trip from x to z can not be broken into shorter legs by making a stopover at some y.)
    (a) Show that the ultrametic property implies the triangle inequality.
    (b) In an ultrametric space show that "all triangles are isosceles."
    (c) Show that a metric space with an ultrametric is totally disconnected.

(d) Define a metric on the set Σ of strings of 0's and 1's in Exercise 104 as

$$d_*(a,b) = \begin{cases} \frac{1}{2^n} & \text{if } n \text{ is the smallest index for which } a_n \neq b_n \\ 0 & \text{if } a = b. \end{cases}$$

Show that d_* is an ultrametric.

(e) Prove that the identity map is a homeomorphism $(\Sigma, d) \to (\Sigma, d_*)$.

*106. $\mathbb{Q}$ inherits the Euclidean metric from $\mathbb{R}$, but it also carries a very different metric, the ***p*-adic** metric. Given a prime number p and an integer n, the p-adic norm of n is

$$|n|_p = \frac{1}{p^k}$$

where p^k is the largest power of p that divides n. (The norm of 0 is by definition 0.) The more factors of p, the smaller the p-norm. Similarly, if $x = a/b$ is a fraction, we factor x as

$$x = p^k \cdot \frac{r}{s}$$

where p divides neither r nor s, and we set

$$|x|_p = \frac{1}{p^k}.$$

The p-adic metric on $\mathbb{Q}$ is

$$d_p(x,y) = |x - y|_p .$$

(a) Prove that d_p is a metric with respect to which $\mathbb{Q}$ is perfect — every point is a cluster point.

(b) Prove that d_p is an ultrametric.

(c) Let $\mathbb{Q}_p$ be the metric space completion of $\mathbb{Q}$ with respect to the metric d_p, and observe that the extension of d_p to $\mathbb{Q}_p$ remains an ultrametric. Infer from Exercise 105 that $\mathbb{Q}_p$ is totally disconnected.

(d) Prove that $\mathbb{Q}_p$ is locally compact, in the sense that every point has small compact neighborhoods.

(e) Infer that $\mathbb{Q}_p$ is covered by neighborhoods homeomorphic to the Cantor set. See Gouvêa's book, *p-adic Numbers*.

107. Prove directly that there is a continuous surjection of the middle-thirds Cantor set C onto the closed interval $[0, 1]$. [Hint: If $x \in C$ has base 3 expansion (x_n) then all the entries are 0's and 2's. Write $y = (y_n)$ by replacing the 2's in (x_n) by 1's and interpreting the answer base 2. Show that the map $x \mapsto y$ works.]

*108. Let P be a closed perfect subset of a separable complete metric space M. Prove that each point of P is a condensation point of P. In symbols, $P = P' \quad \Rightarrow \quad P = P^*$.

109. Let M, N be non-empty metric spaces and $P = M \times N$.

(a) If M, N are perfect, prove that P is perfect.

(b) If M, N are totally disconnected, prove that P is totally disconnected.

(c) What about the converses?

(d) Infer that the Cartesian product of Cantor spaces is a Cantor space. (We already know that the Cartesian product of compacts is compact.)

(e) Why does this imply that $C \times C = \{(x, y) \in \mathbb{R}^2 : x \in C \text{ and } y \in C\}$ is homeomorphic to C, C being the standard Cantor set?

*110. To prove that Antoine's Necklace A is a Cantor set, you need to show that A is compact, perfect, non-empty, and totally disconnected.

(a) Do so. [Hint: What is the diameter of any connected component of A^n, and what does that imply about A?]

(b) If, in the Antoine construction only two linked solid tori are placed inside each larger solid torus, show that the intersection $A = \bigcap A^n$ is not a Cantor set because it is not totally disconnected.

**111. Given a Cantor space $M \subset \mathbb{R}^2$, given a line segment $[p, q] \subset \mathbb{R}^2$ with $p, q \notin M$, and given an $\epsilon > 0$, prove that there exists a path A in the ϵ-neighborhood of $[p, q]$ that joins p to q and is disjoint from M. [Hint: Think of A as a bisector of M. From this bisection fact a dyadic disc partition of M can be constructed, which leads to the proof that M is tame.]

*112. Consider the **Hilbert cube**

$$H = \{(x_1, x_2, \ldots) \in [0,1]^\infty : \text{ for each } n \in \mathbb{N},\ |x_n| \leq 1/2^n\}.$$

Prove that H is compact with respect to the metric

$$d(x, y)) = \sup_n |x_n - y_n|$$

where $x = (x_n)$, $y = (y_n)$. [Hint: sequences of sequences.]
Remark Although compact, H is infinite-dimensional and is homeomorphic to no subset of $\mathbb{R}^m$.

113. Prove that the Hilbert cube is perfect and homeomorphic to its Cartesian square, $H \cong H \times H$.

***114. Assume that M is compact, non-empty, perfect, and homeomorphic to its Cartesian square, $M \cong M \times M$. Must M be homeomorphic to the Cantor set, the Hilbert cube, or some combination of them?

115. A **Peano space** is a metric space M that is the continuous image of the unit interval: there is a continuous surjection $\tau : [0, 1] \to M$. Theorem 68 states the amazing fact that the 2-disc is a Peano space. Prove that every Peano space is
   (a) compact,
   (b) non-empty,
   (c) path-connected,
   *(d) and **locally path-connected**, in the sense that for each $p \in M$ and each neighborhood U of p there is a smaller neighborhood V of p such that any two points of V can be joined by a path in U.

*116. The converse to Exercise 115 is the **Hahn-Mazurkiewicz Theorem**. Assume that a metric space M is a compact, non-empty, path-connected, and locally path-connected. Use the Cantor Surjection Theorem 65 to show that M is a Peano space. [The key is to make uniformly short paths to fill in the gaps of $[0, 1] \setminus C$.]

117. One of the famous theorems in plane topology is the **Jordan Curve Theorem**. It states that if $f : [a, b] \to \mathbb{R}^2$ is continuous, $f(a) = f(b)$, and for no other pair of distinct $s, t \in [a, b]$ does $f(s)$ equal $f(t)$ then the complement of the path f in $\mathbb{R}^2$ consists of two disjoint, connected open sets, its inside and its outside. Prove the Jordan Curve Theorem for the circle, the square, the triangle, and, if you have courage — any simple closed polygon.

118. Prove that there is a continuous surjection $\mathbb{R} \to \mathbb{R}^2$. What about $\mathbb{R}^n$?

119. The **utility problem** gives three houses 1, 2, 3 in the plane and the three utilities, Gas, Water, and Electricity. You are supposed to connect each house to the three utilities without crossing utility lines. (The houses and utilities are disjoint.)

   (a) Use the Jordan curve theorem to show that there is no solution to the utility problem in the plane.
   *(b) Show also that the utility problem cannot be solved on the 2-sphere S^2.
   *(c) Show that the utility problem can be solved on the surface of the torus.
   *(d) What about the surface of the Klein bottle?
   ***(e) Given utilities $U_1, \ldots, U_m$ and houses $H_1, \ldots, H_n$ located on a surface with g handles, find necessary and sufficient conditions on m, n, g so that the utility problem can be solved.

120. The open cylinder is $(0, 1) \times S^1$. The punctured plane is $\mathbb{R}^2 \setminus \{0\}$.
   (a) Prove that the open cylinder is homeomorphic to the punctured plane.
   (b) Prove that the open cylinder, the double cone, and the plane are not homeomorphic.

121. Is the closed strip $\{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1\}$ homeomorphic to the closed halfplane $\{(x, y) \in \mathbb{R}^2 : x \geq 0\}$? Prove or disprove.

122. Is the plane minus four points on the x-axis homeomorphic to the plane minus four points in an arbitrary configuration?

123. Suppose that $A, B \subset \mathbb{R}^2$.
   (a) If A and B are homeomorphic, are their complements homeomorphic?
   *(b) What if A and B are compact?
   ***(c) What if A and B are compact and connected?

**124. Let M be a metric space and let $\mathcal{K}$ denote the class of non-empty compact subsets of M. The r-neighborhood of $A \in \mathcal{K}$ is

$$M_r A = \{x \in M : \exists a \in A \text{ and } d(x, a) < r\} = \bigcup_{a \in A} M_r a.$$

For $A, B \in \mathcal{K}$ define

$$D(A, B) = \inf\{r > 0 : A \subset M_r B \text{ and } B \subset M_r A\}.$$

   (a) Show that D is a metric on $\mathcal{K}$. (It is called the **Hausdorff metric**.)
   (b) Denote by $\mathcal{F}$ the collection of finite non-empty subsets of M and prove that $\mathcal{F}$ is dense in $\mathcal{K}$. That is, given $A \in \mathcal{K}$ and given $\epsilon > 0$ show that there exists $F \in \mathcal{F}$ such that $D(A, F) < \epsilon$.
   (c) If M is compact, prove that $\mathcal{K}$ is compact.
   (d) If M is connected, prove that $\mathcal{K}$ is connected.

***(e) If M is path-connected is $\mathcal{K}$ path-connected?

(f) If M and M' are homeomorphic does it follow that $\mathcal{K}$ and $\mathcal{K}'$ are homeomorphic?

***(g) What about the converse?

125. As on page 105, consider the subsets of $\mathbb{R}$,

$$A = \{0\} \cup [1, 2] \cup \{3\} \quad \text{and} \quad B = \{0\} \cup \{1\} \cup [2, 3].$$

(a) Why is there no ambient homeomorphism of $\mathbb{R}$ to itself that carries A onto B?

(b) Thinking of $\mathbb{R}$ as the x-axis, is there an ambient homeomorphism of $\mathbb{R}^2$ to itself that carries A onto B?

**126. Consider an overhand (trefoil) knot K in $\mathbb{R}^3$. It can be shown that there is no homeomorphism of $\mathbb{R}^3$ to itself that sends K to the standard unit circle $S^1 \subset \mathbb{R}^2$. (See Rolfsen's book, *Knots and Links*.) Thinking of $\mathbb{R}^3$ as the plane $x_4 = 0$ in (x_1, x_2, x_3, x_4)-space $\mathbb{R}^4$, show that there is a homeomorphism of $\mathbb{R}^4$ to itself that carries K onto S^1.

**127. Start with a set $S \subset \mathbb{R}$ and successively take its closure, the complement of its closure, the closure of that, and so on. S, $\mathrm{cl}(S)$, $(\mathrm{cl}(S))^c$, ... Do the same to S^c. In total, how many distinct subsets of $\mathbb{R}$ can be produced this way? In particular decide whether each chain S, $\mathrm{cl}(S)$, ... consists of only finitely many sets. For example, if $S = \mathbb{Q}$ then we get $\mathbb{Q}$, $\mathbb{R}$, $\emptyset$, $\mathbb{R}$, $\emptyset$, $\mathbb{R}$, $\emptyset$, ..., and $\mathbb{Q}^c$, $\mathbb{R}$, $\emptyset$, $\mathbb{R}$, ...for a total of four sets.

**128. Consider the letter T.

(a) Prove that there is no way to place uncountably many copies of the letter T disjointly in the plane. [Hint: First prove this when the unit square replaces the plane.]

(b) Prove that there is no way to place uncountably many homeomorphic copies of the letter T disjointly in the plane.

(c) For which other letters of the alphabet is this true?

(d) Let U be a set in $\mathbb{R}^3$ formed like an umbrella: it is a disc with a perpendicular segment attached to its center. Prove that uncountably many copies of U can not be placed disjointly in $\mathbb{R}^3$.

(e) What if the perpendicular segment is attached to the boundary of the disc?

129. Let M be a complete, separable metric space such as $\mathbb{R}^m$. Prove the **Cupcake Theorem: each closed set $K \subset M$ can be expressed uniquely as the disjoint union of a countable set and a perfect closed set, $C \sqcup P = K$.

*130. Write jingles at least as good as the following.

When a set
in the plane
is closed and bounded,
you can always draw
a curve around it.
Peter Přibik

T'is a most indisputable fact
If you want to make something compact
Make it bounded and closed
For you're totally hosed
If either condition you lack.

Lest the reader infer an untruth
(Which I think would be highly uncouth)
I must hasten to add
There are sets to be had
Where the converse is false, fo'sooth.
Karla Westfahl

A coffee cup feeling quite dazed,
said to a donut, amazed,
an open surjective continuous injection,
You'd be plastic and I'd be glazed.
Norah Esty

If a clopen set can be detected,
Your metric space is disconnected.
David Owens

Pre-lim problems†

1. Suppose that $f : \mathbb{R}^m \to \mathbb{R}$ satisfies two conditions:
   (i) For each compact set K, $f(K)$ is compact.
   (ii) For any nested decreasing sequence of compacts (K_n),

$$f\Big(\bigcap K_n\Big) = \bigcap f(K_n).$$

Prove that f is continuous.

† These are questions taken from the exam given to first year math graduate students at U.C. Berkeley.

2. Let $X \subset \mathbb{R}^m$ be compact and $f : X \to \mathbb{R}$ be continuous. Given $\epsilon > 0$, show that there is a constant M such that for all $x, y \in X$, $|f(x) - f(y)| \leq M\,|x - y| + \epsilon$.
3. Consider $f : \mathbb{R}^2 \to \mathbb{R}$. Assume that for each fixed x_0, $y \mapsto f(x_0, y)$ is continuous and for each fixed y_0, $x \mapsto f(x, y_0)$ is continuous. Find such an f that is not continuous.
4. Let $f : \mathbb{R}^2 \to \mathbb{R}$ satisfy the following properties. For each fixed $x_0 \in \mathbb{R}$ the function $y \mapsto f(x_0, y)$ is continuous and for each fixed $y_0 \in \mathbb{R}$ the function $x \mapsto f(x, y_0)$ is continuous. Also assume that if K is any compact subset of $\mathbb{R}^2$ then $f(K)$ is compact. Prove that f is continuous.
5. Let $f(x, y)$ be a continuous real valued function defined on the unit square $[0, 1] \times [0, 1]$. Prove that

$$g(x) = \max\{f(x, y) : y \in [0, 1]\}$$

is continuous.
6. Let $\{U_k\}$ be a cover of $\mathbb{R}^m$ by open sets. Prove that there is a cover $\{V_k\}$ of $\mathbb{R}^m$ by open sets V_k such that $V_k \subset U_k$ and each compact subset of $\mathbb{R}^m$ is disjoint from all but finitely many of the V_k.
7. A function $f : [0, 1] \to \mathbb{R}$ is said to be **upper semi-continuous** if given $x \in [0, 1]$ and $\epsilon > 0$ there exists a $\delta > 0$ such that $|y - x| < \delta$ implies that $f(y) < f(x) + \epsilon$. Prove that an upper semi-continuous function on $[0, 1]$ is bounded above and attains its maximum value at some point $p \in [0, 1]$.
8. Prove that a continuous function $f : \mathbb{R} \to \mathbb{R}$ which sends open sets to open sets must be monotonic.
9. Show that $[0, 1]$ can not be written as a countably infinite union of disjoint closed subintervals.
10. A **connected component** of a metric space M is a maximal connected subset of M. Give an example of $M \subset \mathbb{R}$ having uncountably many connected components. Can such a subset be open? Closed? Does your answer change if $\mathbb{R}^2$ replaces $\mathbb{R}$?
11. Let $U \subset \mathbb{R}^m$ be an open set. Suppose that the map $h : U \to \mathbb{R}^m$ is a homeomorphism from U onto $\mathbb{R}^m$ which is uniformly continuous. Prove that $U = \mathbb{R}^m$.
12. Let X be a non-empty connected set of real numbers. If every element of X is rational prove that X has only one element.
13. Let $A \subset \mathbb{R}^m$ be compact, $x \in A$. Let (x_n) be a sequence in A such that every convergent subsequence of (x_n) converges to x.
    (a) Prove that the sequence (x_n) converges.
    (b) Give an example to show if A is not compact, the result in (a) is not necessarily true.
14. Assume that $f : \mathbb{R} \to \mathbb{R}$ is uniformly continuous. Prove that there are constants A, B such that $|f(x)| \leq A + B\,|x|$ for all $x \in \mathbb{R}$.
15. Let $h : [0, 1) \to \mathbb{R}$ be a uniformly continuous function where $[0, 1)$ is the half open interval. Prove that there is a unique continuous map $g : [0, 1] \to \mathbb{R}$ such that $g(x) = h(x)$ for all $x \in [0, 1)$.

3

Functions of a Real Variable

1 Differentiation

The function $f : (a, b) \to \mathbb{R}$ is **differentiable at x** if

$$\lim_{t \to x} \frac{f(t) - f(x)}{t - x} = L \tag{1}$$

exists. By definition[†] this means L is a real number and for each $\epsilon > 0$ there exists a $\delta > 0$ such that if $0 < |t - x| < \delta$ then the differential quotient above differs from L by $< \epsilon$. The limit L is the **derivative** of f at x, $f'(x)$. In calculus language, $\Delta x = t - x$ is the change in the independent variable x, while $\Delta f = f(t) - f(x)$ is the resulting change in the dependent variable $y = f(x)$. Differentiability at x means that

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}.$$

We begin by reviewing the proofs of some standard calculus facts.

[†] This concept of limit is slightly different from the limit of a sequence. Here t is a continuous parameter that tends to x, whereas for the sequence (a_n), the parameter n is an integer that grows without bound. A limit definition general enough to include both concepts is discussed in Exercise 26.

1 The Rules of Differentiation

(a) Differentiability implies continuity.

(b) If f and g are differentiable at x then so is $f+g$, the derivative being

$$(f+g)'(x) = f'(x) + g'(x).$$

(c) If f and g are differentiable at x then so is $f \cdot g$, the derivative being

$$(f \cdot g)'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x).$$

(d) The derivative of a constant is zero, $c' = 0$.

(e) If f and g are differentiable at x and $g(x) \neq 0$ then f/g is differentiable at x, the derivative being

$$(f/g)'(x) = (f'(x) \cdot g(x) - f(x) \cdot g'(x))/g(x)^2.$$

(f) If f is differentiable at x and g is differentiable at $y = f(x)$ then $g \circ f$ is differentiable at x, the derivative being

$$(g \circ f)'(x) = g'(y) \cdot f'(x).$$

Proof (a) Continuity in the calculus notation amounts to the assertion that $\Delta f \to 0$ as $\Delta x \to 0$. This is obvious. If the fraction $\Delta f/\Delta x$ tends to a finite limit while its denominator tends to zero, then its numerator must also tend to zero.

(b) Since $\Delta(f+g) = \Delta f + \Delta g$,

$$\frac{\Delta(f+g)}{\Delta x} = \frac{\Delta f}{\Delta x} + \frac{\Delta g}{\Delta x} \to f'(x) + g'(x).$$

as $\Delta x \to 0$.

(c) Since $\Delta(f \cdot g) = \Delta f \cdot g(x+\Delta x) + f(x) \cdot \Delta g$, continuity of g at x implies that

$$\frac{\Delta(f \cdot g)}{\Delta x} = \frac{\Delta f}{\Delta x} g(x+\Delta x) + f(x)\frac{\Delta g}{\Delta x} \to f'(x)g(x) + f(x)g'(x),$$

as $\Delta x \to 0$.

(d) If c is a constant then $\Delta c = 0$ and $c' = 0$.

(e) Since

$$\Delta(f/g) = \frac{g(x)\Delta f - f(x)\Delta g}{g(x+\Delta x)g(x)},$$

the formula follows when we divide by Δx and take the limit.

(f) The shortest proof of the chain rule for $y = f(x)$ is by cancellation:

$$\frac{\Delta g}{\Delta x} = \frac{\Delta g}{\Delta y}\frac{\Delta y}{\Delta x} \to g'(y)f'(x).$$

A slight flaw is present, Δy may be zero when Δx is not. This is not a big problem. Differentiability of g at y implies that

$$\frac{\Delta g}{\Delta y} = g'(y) + \sigma$$

where $\sigma = \sigma(\Delta y) \to 0$ as $\Delta y \to 0$. *Define* $\sigma(0) = 0$. The formula

$$\Delta g = (g'(y) + \sigma)\Delta y$$

holds for all small Δy, including $\Delta y = 0$. Continuity of f at x (which is true by (a)) implies that $\Delta y \to 0$ as $\Delta x \to 0$. Thus

$$\frac{\Delta g}{\Delta x} = (g'(y) + \sigma)\frac{\Delta y}{\Delta x} \to g'(y)f'(x)$$

as $\Delta x \to 0$. □

2 Corollary *The derivative of a polynomial $a_0 + a_1x + \cdots + a_nx^n$ exists everywhere and equals $a_1 + 2a_2x + \cdots + na_nx^{n-1}$.*

Proof Immediate from the differentiation rules. □

A function $f : (a,b) \to \mathbb{R}$ that is differentiable at each $x \in (a,b)$ is **differentiable**.

3 Mean Value Theorem *A continuous function $f : [a,b] \to \mathbb{R}$ that is differentiable on the interval (a,b) has the **mean value property**: there exists a point $\theta \in (a,b)$ such that*

$$f(b) - f(a) = f'(\theta)(b-a).$$

Proof Let

$$S = \frac{f(b) - f(a)}{b-a}$$

be the slope of the secant of the graph of f. See Figure 56.

The function $\phi(x) = f(x) - Sx$ is differentiable and has the same value

$$v = \frac{bf(a) - af(b)}{b-a}$$

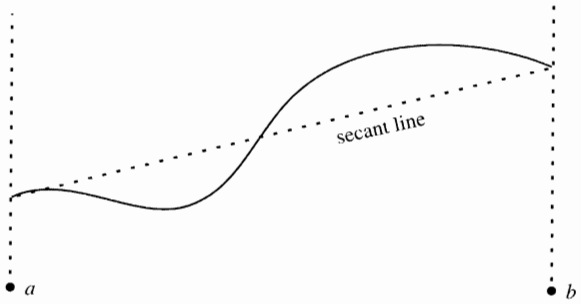


Figure 56 The secant line for the graph of f.

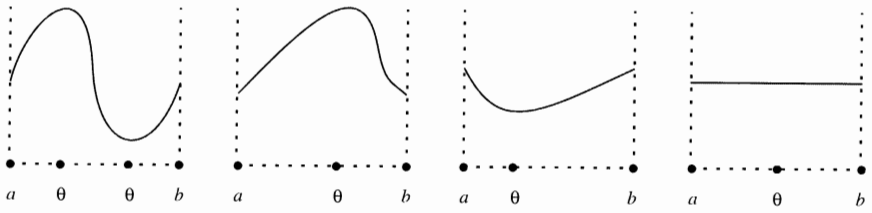


Figure 57 $\phi'(\theta) = 0$.

at a and b. Differentiability implies continuity. ϕ is continuous and therefore takes on maximum and minimum values. Since it has the same value at both endpoints, ϕ has a maximum or a minimum that occurs at a point $\theta \in (a, b)$. See Figure 57. Then $\phi'(\theta) = 0$ (see Exercise 6) and $f(b) - f(a) = f'(\theta)(b - a)$.

4 Corollary *If f is differentiable and $|f'(x)| \leq M$ for all $x \in (a, b)$ then f satisfies the global Lipschitz condition: for all $t, x \in (a, b)$,*

$$|f(t) - f(x)| \leq M\,|t - x|\,.$$

In particular if $f'(x) = 0$ for all $x \in (a, b)$ then $f(x)$ is constant.

Proof $|f(t) - f(x)| = |f'(\theta)(t - x)|$ for some θ between x and t. □

Remark The Mean Value Theorem is the most important theorem in calculus for making estimates.

It is often convenient to deal with two functions simultaneously, and for that we have the following result.

5 Ratio Mean Value Theorem *Suppose that the functions f and g are continuous on an interval $[a, b]$ and differentiable on the interval (a, b).*

Then there is a $\theta \in (a, b)$ such that

$$\Delta f g'(\theta) = \Delta g f'(\theta)$$

where $\Delta f = f(b) - f(a)$ and $\Delta g = g(b) - g(a)$. (If $g(x) \equiv x$, the Ratio Mean Value Theorem becomes the ordinary Mean Value Theorem.)

Proof If $\Delta g \neq 0$, then the theorem states that for some θ,

$$\frac{\Delta f}{\Delta g} = \frac{f'(\theta)}{g'(\theta)}.$$

This ratio expression is how to remember the theorem. The whole point here is that f' and g' are evaluated at the same θ. The function

$$\Phi(x) = \Delta f(g(x) - g(a)) - \Delta g(f(x) - f(a))$$

is differentiable and its value at both endpoints a, b is 0. Since Φ is continuous it takes on a maximum and a minimum somewhere in the interval $[a, b]$. Since Φ has equal values at the endpoints of the interval, it must take on a maximum or minimum at some point $\theta \in (a, b)$; i.e., $\theta \neq a, b$. Then $\Phi'(\theta) = 0$ and $\Delta f g'(\theta) = \Delta g f'(\theta)$ as claimed. □

6 L'Hospital's Rule *If f and g are differentiable functions defined on an interval (a, b), both of which tend to 0 at b, and if the ratio of their derivatives $f'(x)/g'(x)$ tends to a finite limit L at b then $f(x)/g(x)$ also tends to L at b. (We assume that $g(x), g'(x) \neq 0$.)*

Rough proof Let $x \in (a, b)$ tend to b. Imagine a point $t \in (a, b)$ tending to b much faster than x does. It is a kind of "advance guard" for x. Then $f(t)/f(x)$ and $g(t)/g(x)$ are as small as we wish, and by the Ratio Mean Value Theorem, there is a $\theta \in (x, t)$ such that

$$\frac{f(x)}{g(x)} = \frac{f(x) - 0}{g(x) - 0} \doteq \frac{f(x) - f(t)}{g(x) - g(t)} = \frac{f'(\theta)}{g'(\theta)}.$$

The latter tends to L because θ is sandwiched between x and t as they tend to b. The symbol "$\doteq$" means approximately equal. See Figure 58. □

Complete proof Given $\epsilon > 0$ we must find $\delta > 0$ such that if $|x - b| < \delta$ then $|f(x)/g(x) - L| < \epsilon$. Since $f'(x)/g'(x)$ tends to L at b there does exist $\delta > 0$ such that if $x \in (b - \delta, b)$ then

$$\left|\frac{f'(x)}{g'(x)} - L\right| < \frac{\epsilon}{2}.$$

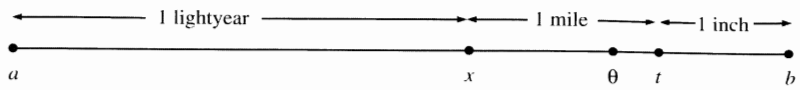


Figure 58 x and t escort θ toward b.

For each $x \in (b-\delta, b)$ determine a point $t \in (b-\delta, b)$ which is so near to b that

$$|f(t)| + |g(t)| < \frac{g(x)^2\epsilon}{4(|f(x)| + |g(x)|)}$$
$$|g(t)| < \frac{|g(x)|}{2}.$$

Since $f(t)$ and $g(t)$ tend to 0 as t tends to b, and since $g(x) \neq 0$ such a t exists. It depends on x, of course. By this choice of t and the Ratio Mean Value Theorem

$$\left|\frac{f(x)}{g(x)} - L\right| = \left|\frac{f(x)}{g(x)} - \frac{f(x)-f(t)}{g(x)-g(t)} + \frac{f(x)-f(t)}{g(x)-g(t)} - L\right|$$
$$\leq \left|\frac{g(x)f(t) - f(x)g(t)}{g(x)(g(x)-g(t))}\right| + \left|\frac{f'(\theta)}{g'(\theta)} - L\right| < \epsilon,$$

which completes the proof that $f(x)/g(x) \to L$ as $x \to b$. □

It is clear that L'Hospital's Rule holds equally well as x tends to b or to a. It is also true that it holds when x tends to $\pm\infty$ or when f and g tend to $\pm\infty$. See Exercises 7, 8.

From now on feel free to use L'Hospital's Rule!

7 Theorem *If f is differentiable on (a, b) then its derivative function $f'(x)$ has the intermediate value property.*

Differentiability of f implies continuity of f, and so the Intermediate Value Theorem applies to f and states that f takes on all intermediate values, but this is not what Theorem 7 is about. Not at all. Theorem 7 concerns f' not f. The function f' can well be discontinuous, but nevertheless it too takes on all intermediate values. In a clear abuse of language, functions like f' possessing the intermediate value property are called **Darboux continuous**, even when they are discontinuous! Darboux was the first to realize how badly discontinuous a derivative function can be. Despite the fact that

f' has the intermediate value property, it can be discontinuous at almost every point of $[a, b]$. Strangely enough, however, f' can not be discontinuous at every point. If f is differentiable, f' must be continuous at a dense, thick set of points. See Exercise 24 and the next section for the relevant definitions.

Proof Suppose that $a < x_1 < x_2 < b$ and

$$\alpha = f'(x_1) < \gamma < f'(x_2) = \beta.$$

We must find $\theta \in (x_1, x_2)$ such that $f'(\theta) = \gamma$.

Choose a small h, $0 < h < x_2 - x_1$, and draw the secant segment $\sigma(x)$ between the points $(x, f(x))$ and $(x+h, f(x+h))$ on the graph of f. Slide x from x_1 to $x_2 - h$ continuously. This is the **sliding secant method**. See Figure 59.

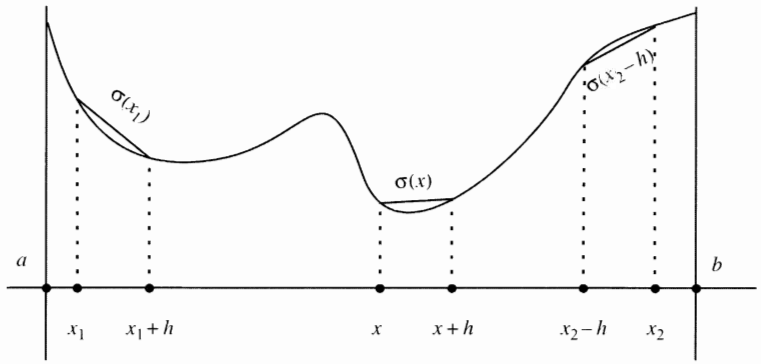


Figure 59 The sliding secant.

When h is small enough, slope $\sigma(x_1) \doteq f'(x_1)$ and slope $\sigma(x_2 - h) \doteq f'(x_2)$. Thus

$$\text{slope}\,\sigma(x_1) < \gamma < \text{slope}\,\sigma(x_2 - h).$$

Continuity of f implies that for some $x \in (x_1, x_2 - h)$, slope $\sigma(x) = \gamma$. The Mean Value Theorem then gives a $\theta \in (x, x+h)$ such that $f'(\theta) = \gamma$. □

8 Corollary *The derivative of a differentiable function never has a jump discontinuity.*

Proof Near a jump, a function omits intermediate values. □

Pathological examples

Non-jump discontinuities of f' may very well occur. The function

$$f(x) = \begin{cases} x^2 \sin \dfrac{1}{x} & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases}$$

is differentiable everywhere, even at $x = 0$, where $f'(0) = 0$. Its derivative function for $x > 0$ is

$$f'(x) = 2x \sin \frac{1}{x} - \cos \frac{1}{x},$$

which oscillates more and more rapidly with amplitude approximately 1 as $x \to 0$. Since $f'(x) \not\to 0$ as $x \to 0$, f' is discontinuous at $x = 0$. Figure 60 shows why f is differentiable at $x = 0$ and has $f'(0) = 0$. Although the graph oscillates wildly at 0, it does so between the envelopes $y = \pm x^2$, and any curve between these envelopes is tangent to the x-axis at the origin. *Study* this example, Figure 60.

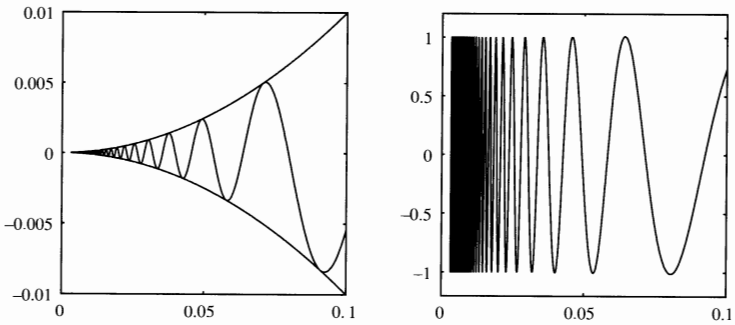


Figure 60 The graphs of the function $y = x^2 \sin(1/x)$ and its envelopes $y = \pm x^2$; and the graph of its derivative.

A similar but worse example is

$$g(x) = \begin{cases} x^{3/2} \sin \dfrac{1}{x} & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases}$$

Its derivative at $x = 0$ is $g'(0) = 0$, while at $x \neq 0$ its derivative is

$$g'(x) = \frac{3}{2}\sqrt{x} \sin \frac{1}{x} - \frac{1}{\sqrt{x}} \cos \frac{1}{x},$$

which oscillates with increasing frequency and unbounded amplitude as $x \to 0$ because $1/\sqrt{x}$ blows up at $x = 0$. See Figure 61.

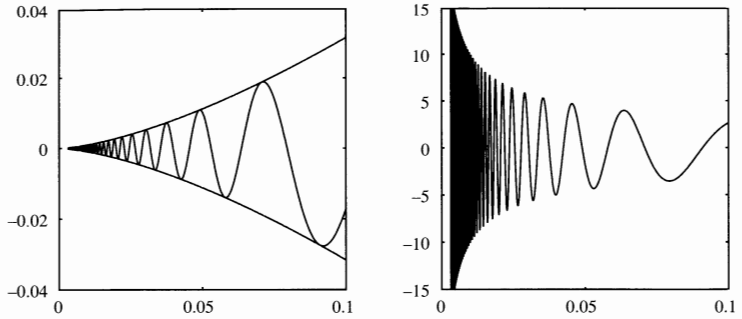


Figure 61 The function $y = x^{3/2} \sin(1/x)$, its envelopes $y = \pm x^{3/2}$, and its derivative.

Higher Derivatives

The derivative of f', if it exists, is the second derivative of f,

$$(f')'(x) = f''(x) = \lim_{t \to x} \frac{f'(t) - f'(x)}{t - x}.$$

Higher derivatives are defined inductively and written $f^{(r)} = (f^{(r-1)})'$. If $f^{(r)}(x)$ exists then f is **r^{th} order differentiable at x**. If $f^{(r)}(x)$ exists for each $x \in (a, b)$ then f is **r^{th} order differentiable**. If $f^{(r)}(x)$ exists for all r and all x then f is **infinitely differentiable**, or **smooth**. The **zero-th derivative** of f is f itself, $f^{(0)}(x) = f(x)$.

9 Theorem *If f is r^{th} order differentiable and $r \ge 1$ then $f^{(r-1)}(x)$ is a continuous function of $x \in (a, b)$.*

Proof Differentiability implies continuity and $f^{(r-1)}(x)$ is differentiable. □

10 Corollary *A smooth function is continuous. Each derivative of a smooth function is smooth and hence continuous.*

Proof Obvious from the definition of smoothness and Theorem 9. □

Smoothness Classes

If f is differentiable and its derivative function $f'(x)$ is a continuous function of x, then f is **continuously differentiable**, and we say that f is of **class C^1**. If f is r^{th} order differentiable and $f^{(r)}(x)$ is a continuous function of x, then f is **continuously r^{th} order differentiable**, and we say that f is

of **class C^r**. If f is smooth, then by the preceding corollary, it is of class C^r for all finite r and we say that f is of **class C^∞**. To round out the notation, we say that a continuous function is of **class C^0**.

Thinking of C^r as the set of functions of class C^r, we have the **regularity hierarchy**,

$$C^0 \supset C^1 \supset \cdots \supset \bigcap_{r\in\mathbb{N}} C^r = C^\infty.$$

Each inclusion $C^r \supset C^{r+1}$ is proper; there exist continuous functions that are not of class C^1, C^1 functions that are not of class C^2, and so on. For example,

$$\begin{aligned} f(x) &= |x| && \text{is of class } C^0 \text{ but not of class } C^1. \\ f(x) &= x\,|x| && \text{is of class } C^1 \text{ but not of class } C^2. \\ f(x) &= |x|^3 && \text{is of class } C^2 \text{ but not of class } C^3 \\ &\cdots \end{aligned}$$

Analytic Functions

A function that can be expressed locally as a convergent power series is **analytic**. More precisely, the function $f : (a, b) \to \mathbb{R}$ is analytic if for each $x \in (a, b)$, there exists a power series

$$\sum a_r h^r$$

and a $\delta > 0$ such that if $|h| < \delta$ then the series converges and

$$f(x + h) = \sum_{r=0}^{\infty} a_r h^r.$$

The concept of series convergence will be discussed further in Section 3 and Chapter 4. Among other things we show in Section 2 of Chapter 4 that analytic functions are smooth; and if $f(x + h) = \sum a_r h^r$ then

$$f^{(r)}(x) = r! a_r.$$

This gives uniqueness of the power series expression of a function: if two power series express the same function f at x then they have identical coefficients, namely $f^{(r)}(x)/r!$. See Exercise 4.36 for a stronger type of uniqueness, namely the identity theorem for analytic functions.

We write C^ω for the class of analytic functions.

A Non-analytic Smooth Function

The fact that smooth functions need not be analytic is somewhat surprising; i.e., C^ω is a proper subset of C^∞. A standard example is

$$e(x) = \begin{cases} e^{-1/x} & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases}$$

Its smoothness is left as an exercise in the use of L'Hospital's rule and induction, Exercise 14. At $x = 0$ the graph of $e(x)$ is infinitely tangent to the x-axis. Every derivative $e^{(r)}(0) = 0$. See Figure 62.

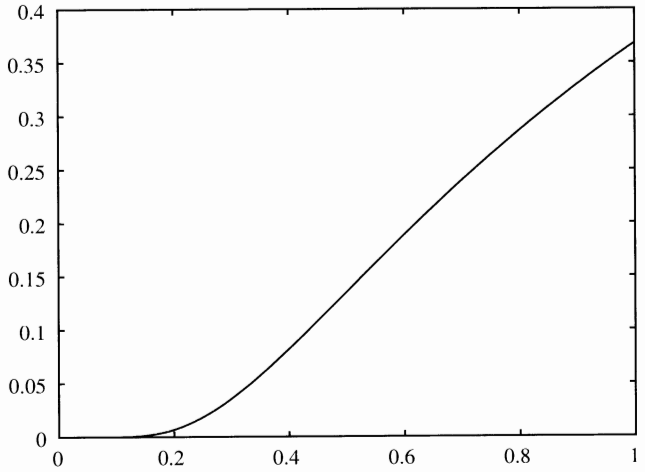


Figure 62 The graph of $e(x) = e^{-1/x}$.

It follows that $e(x)$ is not analytic. For if it were then it could be expressed near $x = 0$ as a convergent series $e(h) = \sum a_r h^r$, and $a_r = e^{(r)}(0)/r!$. Thus $a_r = 0$ for each r, and the series converges to zero, whereas $e(h)$ is different from zero when $h > 0$. Although not analytic at $x = 0$, $e(x)$ is analytic elsewhere. See also Exercise 4.35.

Taylor Approximation

The r^{th} order **Taylor polynomial** of an r^{th} order differentiable function f at x is

$$P(h) = f(x) + f'(x)h + \frac{f''(x)}{2!}h^2 + \cdots + \frac{f^{(r)}(x)}{r!}h^r = \sum_{k=0}^{r} \frac{f^{(k)}(x)}{k!}h^k.$$

The coefficients $f^{(k)}(x)/k!$ are constants, the variable is h. Differentiation of P with respect to h at $h = 0$ gives

$$\begin{aligned} P(0) &= f(x) \\ P'(0) &= f'(x) \\ &\cdots \\ P^{(r)}(0) &= f^{(r)}(x) \end{aligned}$$

11 Taylor Approximation Theorem *Assume that $f : (a, b) \to \mathbb{R}$ is r^{th} order differentiable at x. Then*

(a) P approximates f to order r at x in the sense that the Taylor remainder

$$R(h) = f(x + h) - P(h)$$

is r^{th} order flat at $h = 0$; i.e., $R(h)/h^r \to 0$ as $h \to 0$.

(b) The Taylor polynomial is the only polynomial of degree $\leq r$ with this approximation property.

(c) If, in addition, f is $(r + 1)^{st}$ order differentiable on (a, b) then for some θ between x and $x + h$,

$$R(h) = \frac{f^{(r+1)}(\theta)}{(r+1)!}h^{r+1}.$$

Proof (a) The first r derivatives of $R(h)$ exist and equal 0 at $h = 0$. If $h > 0$ then repeated applications of the Mean Value Theorem give

$$\begin{aligned} R(h) = R(h) - 0 = R'(\theta_1)h = (R'(\theta_1) - 0)h = R''(\theta_2)\theta_1 h \\ = \cdots = R^{(r-1)}(\theta_{r-1})\theta_{r-2} \ldots \theta_1 h \end{aligned}$$

where $0 < \theta_{r-1} < \cdots < \theta_1 < h$. Thus

$$\left|\frac{R(h)}{h^r}\right| = \left|\frac{R^{(r-1)}(\theta_{r-1})\theta_{r-2} \ldots \theta_1 h}{h^r}\right| \leq \left|\frac{R^{(r-1)}(\theta_{r-1}) - 0}{\theta_{r-1}}\right| \to 0$$

as $h \to 0$. If $h < 0$ the same is true with $h < \theta_1 < \cdots < \theta_{r-1} < 0$.

(b) If $Q(h)$ is a polynomial of degree $\leq r$, $Q \neq P$, then $Q - P$ is not r^{th} order flat at $h = 0$, so $f(x + h) - Q(h)$ can not be r^{th} order flat either.

(c) Fix $h > 0$ and define

$$g(t) = f(x + t) - P(t) - \frac{R(h)}{h^{r+1}}t^{r+1} = R(t) - R(h)\frac{t^{r+1}}{h^{r+1}}$$

for $0 \leq t \leq h$. Note that since $P(t)$ is a polynomial of degree r, $P^{(r+1)}(t) = 0$ for all t, and

$$g^{(r+1)}(t) = f^{(r+1)}(x + t) - (r + 1)!\frac{R(h)}{h^{r+1}}.$$

Also, $g(0) = g'(0) = \cdots = g^{(r)}(0) = 0$, and $g(h) = R(h) - R(h) = 0$. Since $g = 0$ at 0 and h, the Mean Value Theorem gives a $t_1 \in (0, h)$ such that $g'(t_1) = 0$. Since $g'(0)$ and $g'(t_1) = 0$, the Mean Value Theorem gives a $t_2 \in (0, t_1)$ such that $g''(t_2) = 0$. Continuing, we get a sequence $t_1 > t_2 > \cdots > t_{r+1} > 0$ such that $g^{(k)}(t_k) = 0$. The $(r + 1)^{\text{st}}$ equation, $g^{(r+1)}(t_{r+1}) = 0$, implies that

$$0 = f^{(r+1)}(x + t_{r+1}) - (r + 1)!\frac{R(h)}{h^{r+1}}.$$

Thus, $\theta = x + t_{r+1}$ makes the equation in (c) true. If $h < 0$ the argument is symmetric. □

12 Corollary *For each $r \in \mathbb{N}$, the smooth non-analytic function $e(x)$ satisfies $\lim_{h\to 0} e(h)/h^r = 0$.*

Proof Obvious from the theorem and the fact that $e^{(r)}(0) = 0$ for all r. □

The **Taylor series** at x of a smooth function f is the infinite Taylor polynomial

$$T(h) = \sum_{r=0}^{\infty} \frac{f^r(x)}{r!}h^r.$$

In calculus, you compute the Taylor series of functions such as $\sin x$, $\arctan x$, e^x, etc. These functions are analytic: their Taylor series converge and express them as power series. In general, however, the Taylor series of a smooth function need not converge to the function, and in fact it may fail to converge at all. The function $e(x)$ is an example of the first phenomenon. Its Taylor series at $x = 0$ converges, but gives the wrong answer. Examples of divergent and totally divergent Taylor series are indicated in Exercise 4.35.

The convergence of a Taylor series is related to how quickly the r^{th} derivative grows (in magnitude) as $r \to \infty$. In Section 6 of Chapter 4 we give necessary and sufficient conditions on the growth rate that determine whether a smooth function is analytic.

Inverse Functions

A strictly monotone continuous function $f : (a, b) \to \mathbb{R}$ bijects (a, b) onto some interval (c, d) where $c = f(a)$, $d = f(b)$ in the increasing case. It is a homeomorphism $(a, b) \to (c, d)$ and its inverse function $f^{-1} : (c, d) \to (a, b)$ is also a homeomorphism. These facts were proved in Chapter 2.

Does differentiability of f imply differentiability of f^{-1}? If $f' \neq 0$ the answer is "yes." Keep in mind, however, the function $f : x \mapsto x^3$. It shows that differentiability of f^{-1} fails when $f'(x) = 0$. For the inverse function is $y \mapsto y^{1/3}$, which is non-differentiable at $y = 0$.

13 Inverse Function Theorem in dimension one *If $f : (a, b) \to (c, d)$ is a differentiable surjection and $f'(x)$ is never zero then f is a homeomorphism, and its inverse is differentiable with derivative*

$$(f^{-1})'(y) = \frac{1}{f'(x)}$$

where $y = f(x)$.

Proof If f' is never zero then by the intermediate value property of derivatives, it is either always positive or always negative. We assume for all x that $f'(x) > 0$. If $a < s < t < b$ then by the Mean Value Theorem there exists $\theta \in (s, t)$ such that $f(t) - f(s) = f'(\theta)(t - s) > 0$. Thus f is strictly monotone. Differentiability implies continuity, so f is a homeomorphism $(a, b) \to (c, d)$. To check differentiability of f^{-1} at $y \in (c, d)$, define

$$x = f^{-1}(y) \quad \text{and} \quad \Delta x = f^{-1}(y + \Delta y) - x.$$

Then $y = f(x)$ and $\Delta y = f(x + \Delta x) - x = \Delta f$. Thus

$$\frac{\Delta f^{-1}}{\Delta y} = \frac{f^{-1}(y + \Delta y) - f^{-1}(y)}{\Delta y} = \frac{\Delta x}{\Delta y} = \frac{1}{\Delta y / \Delta x} = \frac{1}{\Delta f / \Delta x}.$$

Since f is a homeomorphism, $\Delta x \to 0$ if and only if $\Delta y \to 0$, so the limit of $\Delta f^{-1}/\Delta y$ exists and equals $1/f'(x)$. □

If a homeomorphism f and its inverse are both of class C^r, $r \geq 1$, then f is a **C^r diffeomorphism**.

14 Corollary *If $f : (a, b) \to (c, d)$ is a homeomorphism of class C^r, $1 \leq r \leq \infty$, and $f' \neq 0$ then f is a C^r diffeomorphism.*

Proof If $r = 1$, the formula $(f^{-1})'(y) = 1/f'(x) = 1/f'(f^{-1}(y))$ implies that $(f^{-1})'(y)$ is continuous, so f is a C^1 diffeomorphism. Induction on $r \geq 2$ completes the proof. □

The corollary remains true for analytic functions: the inverse of an analytic function with non-vanishing derivative is analytic. The generalization of the inverse function theorem to higher dimensions is a principal goal of Chapter 5.

A longer but more geometric proof of the one dimensional inverse function theorem can be done in two steps.

(i) A function is differentiable if and only if its graph is differentiable.

(ii) The graph of f^{-1} is the reflection of the graph of f across the diagonal, and is thus differentiable.

See Figure 63.

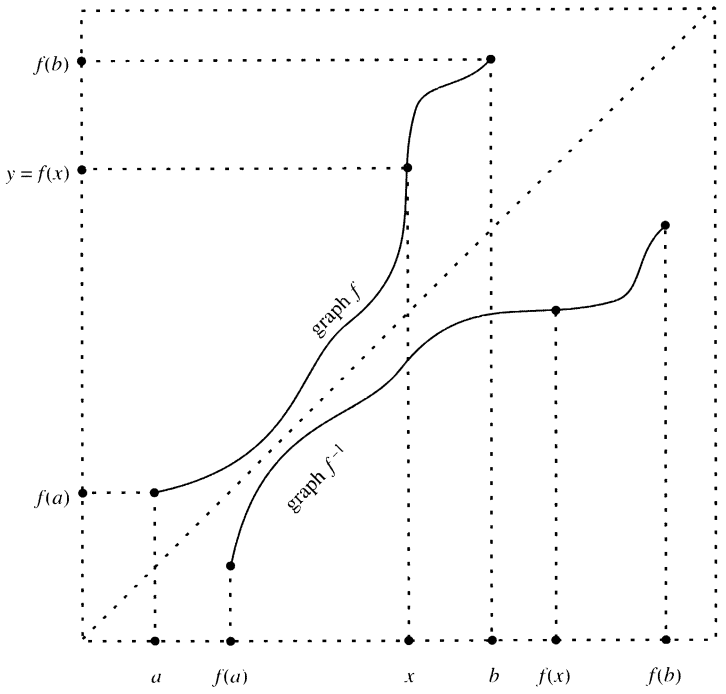


Figure 63 A picture proof of the inverse function theorem in $\mathbb{R}$.

2 Riemann Integration

Let $f : [a, b] \to \mathbb{R}$ be given. Intuitively, the integral of f is the area under its graph; i.e., for $f \geq 0$,

$$\int_a^b f(x)\, ds = \text{area } \mathbb{U}$$

where $\mathbb{U}$ is the **undergraph** of f,

$$\mathbb{U} = \{(x, y) : a \leq x \leq b \text{ and } 0 \leq y \leq f(x)\}.$$

The precise definition involves approximation. A **partition pair** consists of two finite sets of points $P, T \subset [a, b]$; $P = \{x_0, \ldots, x_n\}$ and $T = \{t_1, \ldots, t_n\}$, interlaced as

$$a = x_0 \leq t_1 \leq x_1 \leq t_2 \leq x_2 \leq \cdots \leq t_n \leq x_n = b.$$

We assume the points $x_0, \ldots, x_n$ are distinct. The **Riemann sum** corresponding to f, P, T is

$$R(f, P, T) = \sum_{i=1}^{n} f(t_i)\Delta x_i = f(t_1)\Delta x_1 + f(t_2)\Delta x_2 + \cdots + f(t_n)\Delta x_n$$

where $\Delta x_i = x_i - x_{i-1}$. The Riemann sum R is the area of rectangles which approximate the area under the graph of f. See Figure 64. Think of the points t_i as "sample points." We sample the value of the function f at t_i.

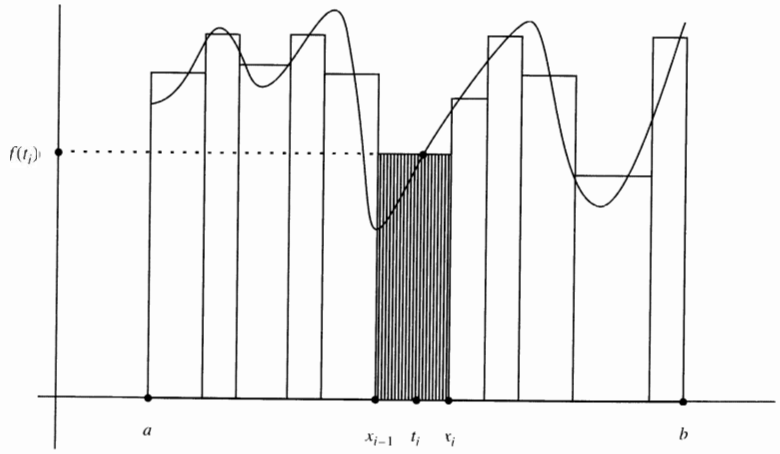


Figure 64 The area of the strip is $f(t_i)\Delta x_i$.

The **mesh** of the partition P is the length of the largest subinterval $[x_{i-1}, x_i]$. A partition with large mesh is **coarse**; one with small mesh is **fine**. In general, the finer the better. A real number I is the **Riemann integral** of f over $[a, b]$ if it satisfies the following approximation condition:

$$\forall \epsilon > 0 \ \exists \delta > 0 \text{ such that if } P, T \text{ is any partition pair then}$$
$$\text{mesh } P < \delta \quad \Rightarrow \quad |R - I| < \epsilon$$

where $R = R(f, P, T)$. If such an I exists it is unique, we denote it as

$$\int_a^b f(x)\, dx = I = \lim_{\text{mesh } P \to 0} R(f, P, T),$$

and we say that f is **Riemann integrable** with Riemann integral I. See Exercise 26 for a formalization of this limit definition.

15 Theorem *If f is Riemann integrable then it is bounded.*

Proof Suppose not. Let $I = \int_a^b f(x)\, dx$. There is some $\delta > 0$ such that $|R - I| < 1$ for all partition pairs P, T with mesh $P < \delta$. Fix such a partition pair $P = \{x_0, \ldots, x_n\}$, $T = \{t_1, \ldots, t_n\}$. If f is unbounded on $[a, b]$ then there is also a subinterval $[x_{i_0-1}, x_{i_0}]$ on which it is unbounded. Choose a new set $T' = \{t_1', \ldots, t_n'\}$ with $t_i' = t_i$ for all $i \neq i_0$ and choose t_{i_0}' such that

$$|f(t_{i_0}') - f(t_{i_0})|\Delta x_{i_0} > 2.$$

This is possible since the supremum of $\{|f(t)| : x_{i_0-1} \leq t \leq x_{i_0}\}$ is ∞. Let $R' = R(f, P, T')$. Then $|R - R'| > 2$, contrary to the fact that both R and R' differ from I by < 1. □

Let $\mathcal{R}$ denote the set of all functions that are Riemann integrable over $[a, b]$.

16 Theorem (Linearity of the Integral)

(a) $\mathcal{R}$ is a vector space and $f \mapsto \int_a^b f(x)\, dx$ is a linear map $\mathcal{R} \to \mathbb{R}$.

(b) The constant function $h(x) = k$ is integrable and its integral is $k(b - a)$.

Proof (a) Riemann sums behave naturally under linear combination:

$$R(f + cg, P, T) = R(f, P, T) + cR(g, P, T),$$

and it follows that their limits, as mesh $P \to 0$, give the expected formula

$$\int_a^b f(x) + cg(x)\, dx = \int_a^b f(x)\, dx + c\int_a^b g(x)\, dx.$$

(b) Every Riemann sum for the constant function $h(x) = k$ is $k(b-a)$, so its integral equals this number too. □

17 Theorem (Monotonicity of the Integral) *If $f, g \in \mathcal{R}$ and $f \le g$ then*

$$\int_a^b f(x)\,dx \le \int_a^b g(x)\,dx.$$

Proof For each partition pair P, T, we have $R(f, P, T) \le R(g, P, T)$. □

18 Corollary *If $f \in \mathcal{R}$ and $|f| \le M$ then $\left|\int_a^b f(x)\,dx\right| \le M(b-a)$.*

Proof By Theorem 16, the constant functions $\pm M$ are integrable. By Theorem 17, $-M \le f(x) \le M$ implies that

$$-M(b-a) \le \int_a^b f(x)\,dx \le M(b-a).$$

Darboux integrability

The **lower sum** and **upper sum** of a function $f : [a, b] \to [-M, M]$ with respect to a partition P of $[a, b]$ are

$$L(f, P) = \sum_{i=1}^{n} m_i \Delta x_i \quad \text{and} \quad U(f, P) = \sum_{i=1}^{n} M_i \Delta x_i$$

where

$$m_i = \inf\{f(t) : x_{i-1} \le t \le x_i\},$$
$$M_i = \sup\{f(t) : x_{i-1} \le t \le x_i\}.$$

We assume f is bounded in order to be sure that m_i and M_i are real numbers. Clearly

$$L(f, P) \le R(f, P, T) \le U(f, P)$$

for all partition pairs P, T. See Figure 65.

The **lower integral** and **upper integral** of f over $[a, b]$ are

$$\underline{I} = \sup_P L(f, P) \quad \text{and} \quad \overline{I} = \inf_P U(f, P).$$

P ranges over all partitions of $[a, b]$ when we take the supremum and infimum. If the lower and upper integrals of f are equal, $\underline{I} = \overline{I}$, then f is **Darboux integrable** and I is its **Darboux integral**.

19 Theorem *Riemann integrability is equivalent to Darboux integrability, and when a function is integrable, its three integrals — lower, upper, and Riemann — are equal.*

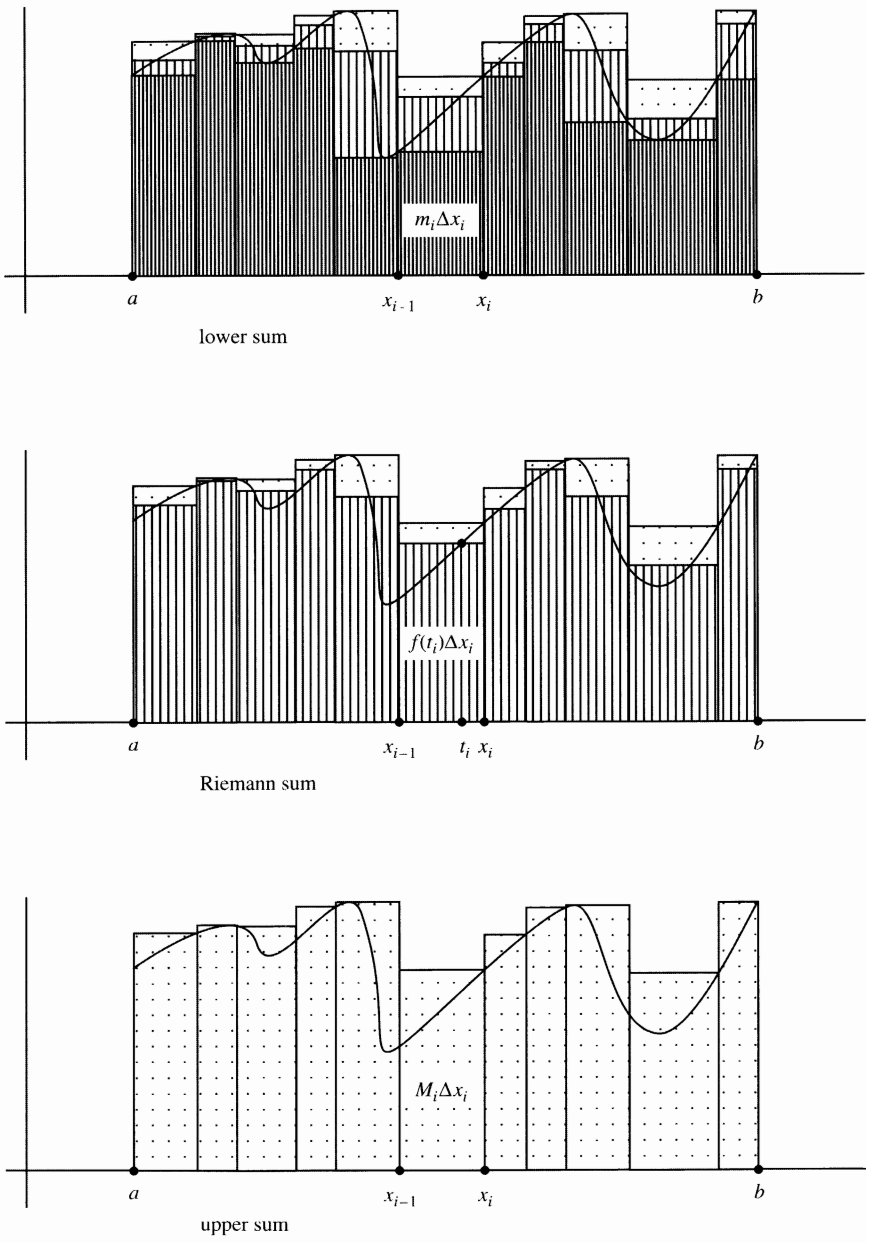


Figure 65 The lower sum, the Riemann sum, and the upper sum.

To prove Theorem 19 it is convenient to refine a partition P by adding more partition points; the partition P' **refines** P if $P' \supset P$.

Suppose first that $P' = P \cup \{w\}$ where $w \in (x_{i_0-1}, x_{i_0})$. The lower sums for P and P' are the same except that $m_{i_0}\Delta x_{i_0}$ in $L(f, P)$ splits into two terms in $L(f, P')$. The sum of the two terms is at least as large as $m_{i_0}\Delta x_{i_0}$. For the infimum of f over the intervals $[x_{i_0-1}, w]$ and $[w, x_{i_0}]$ is at least as large as m_{i_0}. Similarly, $U(f, P') \le U(f, P)$. See Figure 66.

Repetition continues the pattern and we formalize it as the

Refinement Principle *Refining a partition causes the lower sum to increase and the upper sum to decrease.*

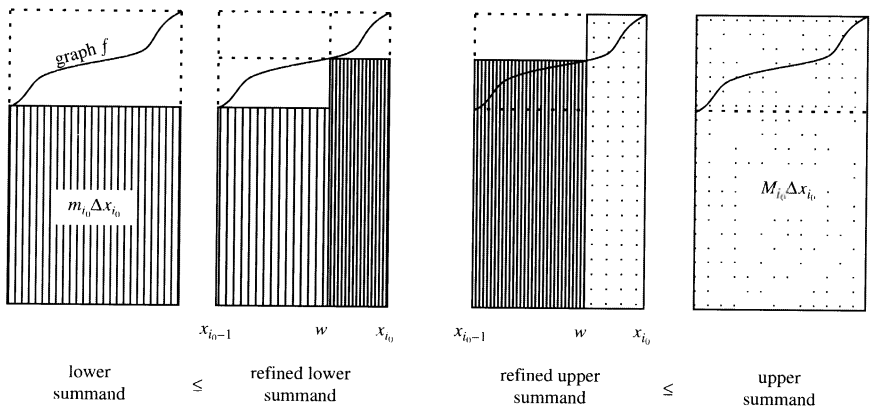


Figure 66 Refinement increases L and decreases U.

The **common refinement** P^* of two partitions P, P' of $[a, b]$ is

$$P^* = P \cup P'.$$

According to the refinement principle

$$L(f, P) \le L(f, P^*) \le U(f, P^*) \le U(f, P').$$

We conclude: each lower sum is less than or equal to each upper sum; the lower integral is less than or equal to the upper; and thus

(2) A bounded function $f : [a, b] \to \mathbb{R}$ is Darboux integrable if and only if $\forall \epsilon > 0 \ \exists P$ such that $U(f, P) - L(f, P) < \epsilon$.

Proof of Theorem 19 Assume that f is Darboux integrable: the lower and upper integrals are equal, say their common value is I. Given $\epsilon > 0$ we must find $\delta > 0$ such that $|R - I| < \epsilon$ whenever $R = R(f, P, T)$ is a

Riemann sum with mesh $P < \delta$. By Darboux integrability and (2) there is a partition P_1 of $[a, b]$ such that

$$U_1 - L_1 < \frac{\epsilon}{2}$$

where $U_1 = U(f, P_1)$, $L_1 = L(f, P_1)$. Fix $\delta = \epsilon/8Mn_1$ where n_1 is the number of partition points in P_1. Let P be any partition with mesh $P < \delta$. (Since $\delta \ll \epsilon$, think of P_1 as coarse and P as much finer.) Let P^* be the common refinement $P \cup P_1$. By the refinement principle,

$$L_1 \le L^* \le U^* \le U_1$$

where $L^* = L(f, P^*)$ and $U^* = U(f, P^*)$. Thus,

$$U^* - L^* < \frac{\epsilon}{2}.$$

Write $P = \{x_i\}$ and $P^* = \{x_j^*\}$ for $0 \le i \le n$ and $0 \le j \le n^*$. The sums $U = \sum M_i \Delta x_i$ and $U^* = \Sigma M_j^* \Delta x_j^*$ are identical except for terms with

$$x_{i-1} < x_j^* < x_i$$

for some i, j. There are at most $n_1 - 2$ such terms and each is of magnitude at most $M\delta$. Thus,

$$U - U^* < (n_1 - 2)2M\delta < \frac{\epsilon}{4}.$$

Similarly, $L^* - L < \epsilon/4$, and so

$$U - L = (U - U^*) + (U^* - L^*) + (L^* - L) < \epsilon.$$

Since I and R both belong to the interval $[L, U]$, we see that $|R - I| < \epsilon$. Therefore f is Riemann integrable.

Conversely, assume that f is Riemann integrable with Riemann integral I. By Theorem 15, f is bounded. Let $\epsilon > 0$ be given. There exists a $\delta > 0$ such that for all partition pairs P, T with mesh $P < \delta$, $|R - I| < \epsilon/4$ where $R = R(f, P, T)$. Fix any such P and consider $L = L(f, P)$, $U = U(f, P)$. There are choices of intermediate sets $T = \{t_i\}$, $T' = \{t_i'\}$ such that each $f(t_i)$ is so close to m_i and each $f(t_i')$ is so close to M_i that $R - L < \epsilon/4$ and $U - R' < \epsilon/4$ where $R = R(f, P, T)$ and $R' = R(f, P, T')$. Since mesh $P < \delta$, we know that $|R - I| < \epsilon/4$ and $|R' - I| < \epsilon/4$. Thus

$$U - L = (U - R') + (R' - I) + (I - R) + (R - L) < \epsilon.$$

Since $\underline{I}$, I, $\overline{I}$ are fixed numbers that belong to the interval $[L, U]$ of length ϵ, and ϵ is arbitrary, the ϵ-principle implies that

$$\underline{I} = I = \overline{I},$$

which proves that f is Darboux integrable and that its lower, upper, and Riemann integrals are equal. □

According to Theorem 19 and (2) we get

20 Riemann's Integrability Criterion

A bounded function is Riemann integrable if and only if
$\forall \epsilon > 0 \; \exists P$ *such that* $U(f, P) - L(f, P) < \epsilon$.

Example Every continuous function $f : [a, b] \to \mathbb{R}$ is Riemann integrable. (See also Corollary 22 to the Riemann-Lebesgue Theorem, below.) Since $[a, b]$ is compact and f is continuous, f is uniformly continuous. See Theorem 44 in Chapter 2. Let $\epsilon > 0$ be given. Uniform continuity provides a $\delta > 0$ such that if $|t - s| < \delta$ then $|f(t) - f(s)| < \epsilon/2(b - a)$. Choose any partition P with mesh $P < \delta$. On each partition interval $[x_{i-1}, x_i]$, we have $M_i - m_i < \epsilon/(b - a)$. Thus

$$U - L = \sum_{i=1}^{n} (M_i - m_i)\Delta x_i < \frac{\epsilon}{(b-a)} \sum \Delta x_i = \epsilon.$$

By Riemann's integrability criterion f is Riemann integrable.

Example A **piecewise continuous function** is continuous except at a finite number of points. A **step function** is constant except at a finite number of points where it is discontinuous. Clearly, a step function is a special type of piecewise continuous function. See Figure 67.

The **characteristic function** (or **indicator function**) of a set $E \subset \mathbb{R}$, χ_E, is 1 at points of E and 0 at points of E^c. See Figure 68.

A step function is a finite sum of constants times characteristic functions of intervals. See Figure 67. Bounded piecewise continuous functions are Riemann integrable. See Corollary 23 below. Some characteristic functions are Riemann integrable, others aren't.

Example The characteristic function of $\mathbb{Q}$ is not integrable on $[a, b]$. It is defined as $\chi_{\mathbb{Q}}(x) = 1$ when $x \in \mathbb{Q}$ and $\chi_{\mathbb{Q}}(x) = 0$ when $x \notin \mathbb{Q}$. See Figure 69. Every lower sum $L(\chi_{\mathbb{Q}}, P)$ is 0 and every upper sum is $b - a$. By Riemann's criterion, $\chi_{\mathbb{Q}}$ is not integrable. Note that $\chi_{\mathbb{Q}}$ is discontinuous at every point, not merely at rational points.

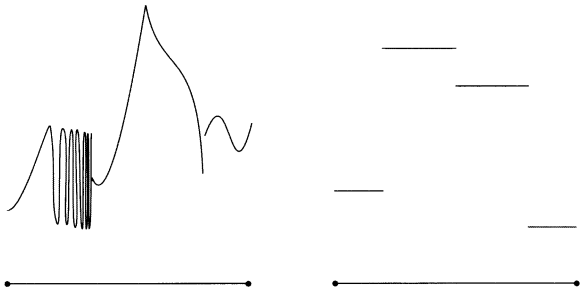


Figure 67 The graphs of a piecewise continuous function and a step function.

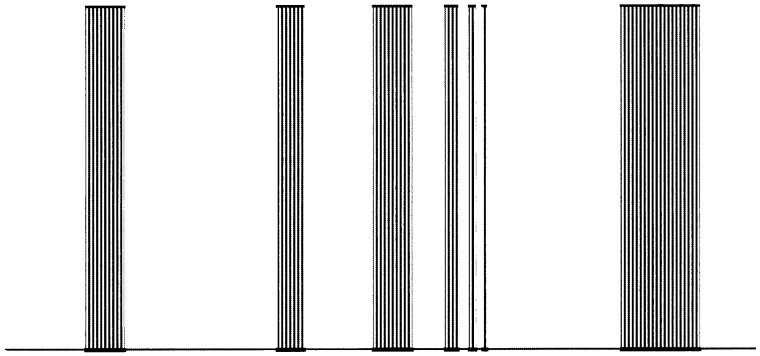


Figure 68 The graph of a characteristic function and the region below the graph.

The fact that $\chi_{\mathbb{Q}}$ fails to be Riemann integrable is actually a failing of Riemann integration theory, for the function $\chi_{\mathbb{Q}}$ is fairly tame. Its integral ought to exist and it ought to be 0, because the undergraph is just countably many line segments of height 1, and their area ought to be 0.

Example The **rational ruler function** is Riemann integrable. At each rational number $x = p/q$, we set $f(x) = 1/q$, while $f(x) = 0$ when x is irrational. See Figure 70. The integral of f is 0. Note that f is discontinuous at every $x \in \mathbb{Q}$ and is continuous at every $x \in \mathbb{Q}^c$.

Example The **Zeno's staircase function** $Z(x) = 1/2$ on the first half of $[0, 1]$, $Z(x) = 3/4$ on the next quarter of $[0, 1]$, and so on. See Figure 71. It is Riemann integrable and its integral is $2/3$. The function has infinitely

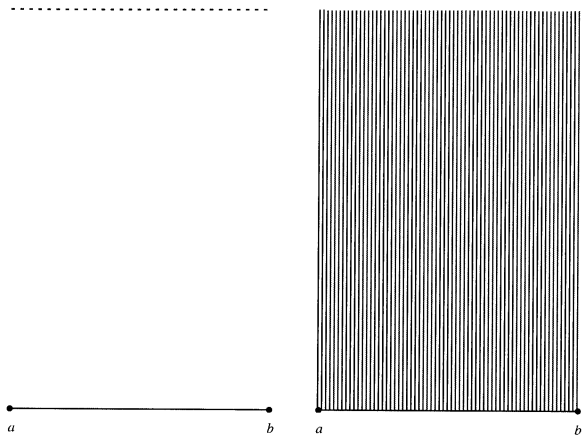


Figure 69 The graph of $\chi_{\mathbb{Q}}$ and the region below its graph.

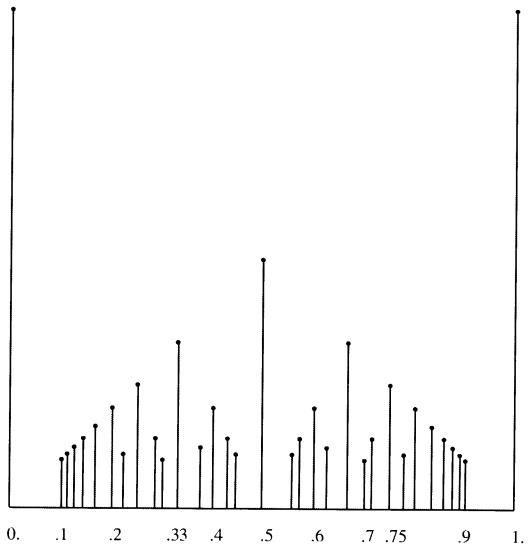


Figure 70 The graph of the rational ruler function and the region below its graph.

many discontinuity points, one at each point $(2^k - 1)/2^k$. In fact, every monotone function is Riemann integrable.[†] See Corollary 24 below.

[†] To prove this directly is not hard. See also Corollary 24 below. The key observation to make is that a monotone function is not much different than a continuous function. It has only jump discontinuities,

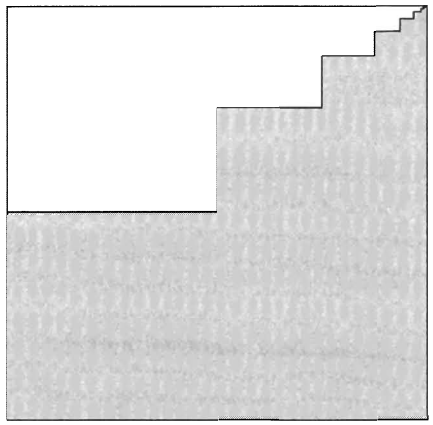


Figure 71 Zeno's staircase.

These examples raise a natural question:

Exactly which functions are Riemann integrable?

To give an answer to the question, and for many other applications, the following concept is very useful. A set $Z \subset \mathbb{R}$ is a **zero set** if for each $\epsilon > 0$ there is a countable covering of Z by open intervals (a_i, b_i) such that

$$\sum_{i=1}^{\infty} b_i - a_i \leq \epsilon.$$

The sum of the series is the **total length** of the covering. Think of zero sets as negligible; if a property holds for all points except those in a zero set then one says that the property holds **almost everywhere**, abbreviated "a.e."

21 Riemann-Lebesgue Theorem *A function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if it is bounded and its set of discontinuity points is a zero set.*

The set D of discontinuity points is exactly what its name implies,

$$D = \{x \in [a, b] : f \text{ is discontinuous at the point } x\}.$$

A function whose set of discontinuity points is a zero set is continuous almost everywhere. The Riemann-Lebesgue theorem states that a function is Riemann integrable if and only if it is bounded and continuous almost everywhere.

Examples of zero sets are

and only countably many of them; given any $\epsilon > 0$, there are only finitely many at which the jump is $\geq \epsilon$. See Exercise 2.30.

(a) Any subset of a zero set.
(b) Any finite set.
(c) Any countable union of zero sets.
(d) Any countable set.
(e) The middle-thirds Cantor set.

(a) is clear. For if $Z_0 \subset Z$ where Z is a zero set, and if $\epsilon > 0$ is given, then there is some open covering of Z by intervals whose total length is $\leq \epsilon$; but the same collection of intervals covers Z_0, which shows that Z_0 is also a zero set.

(b) Let $Z = \{z_1, \ldots, z_n\}$ be a finite set and let $\epsilon > 0$ be given. The intervals $(z_i - \epsilon/2n,\ z_i + \epsilon/2n)$, for $i = 1, \ldots, n$, cover Z and have total length ϵ. Therefore Z is a zero set. In particular, any single point is a zero set.

(c) This is a typical "**$\epsilon/2^n$-argument**." Let $Z_1, Z_2, \ldots$ be a sequence of zero sets and $Z = \bigcup Z_j$. We claim that Z is a zero set. Let $\epsilon > 0$ be given. The set Z_1 can be covered by countably many intervals (a_{i1}, b_{i1}) with total length $\sum(b_{i1} - a_{i1}) \leq \epsilon/2$. The set Z_2 can be covered by countably many intervals (a_{i2}, b_{i2}) with total length $\sum(b_{i2} - a_{i2}) \leq \epsilon/4$. In general, the set Z_j can be covered by countably many intervals (a_{ij}, b_{ij}) with total length

$$\sum_{i=1}^{\infty}(b_{ij} - a_{ij}) \leq \frac{\epsilon}{2^j}.$$

Since the countable union of countable sets is countable, the collection of all the intervals (a_{ij}, b_{ij}) is a countable covering of Z by open intervals, and the total length of all these intervals is

$$\sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty} b_{ij} - a_{ij}\right) \quad \leq \quad \sum_{j=1}^{\infty}\frac{\epsilon}{2^j} \quad = \quad \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{8} + \cdots = \epsilon.$$

Thus Z is a zero set and (c) is proved.

(d) This is implied by (b) and (c).

(e) Let $\epsilon > 0$ be given and choose $n \in \mathbb{N}$ such that $2^n/3^n < \epsilon$. The middle-thirds Cantor set C is contained inside 2^n closed intervals of length $1/3^n$, say $I_1, \ldots, I_{2^n}$. Enlarge each closed interval I_i to an open interval $(a_i, b_i) \supset I_i$ such that $b_i - a_i = \epsilon/2^n$. (Since $1/3^n < \epsilon/2^n$, and I_i has length $1/3^n$, this is possible.) The total length of these 2^n intervals (a_i, b_i) is ϵ. Thus C is a zero set.

In the proof of the Riemann-Lebesgue Theorem, it is useful to focus on the "size" of a discontinuity. A simple expression for this size is the

oscillation of f at x,

$$\operatorname{osc}_x(f) = \limsup_{t\to x} f(t) - \liminf_{t\to x} f(t).$$

Equivalently,

$$\operatorname{osc}_x(f) = \lim_{r\to 0} \operatorname{diam} f([x - r,\ x + r])$$

(Of course, $r > 0$.) It is clear that f is continuous at x if and only if $\operatorname{osc}_x(f) = 0$. It is also clear that if I is any interval containing x in its interior then

$$M_I - m_I \geq \operatorname{osc}_x(f)$$

where M_I and m_I are the supremum and infimum of $f(t)$ as t varies in I. See Figure 72.

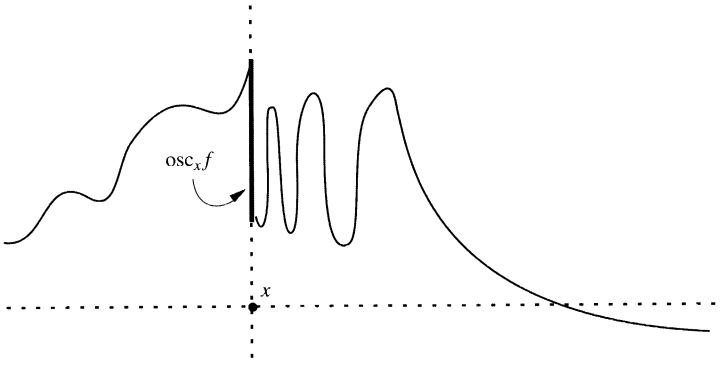


Figure 72 The oscillation of f at x is $\limsup_{t\to x} f(t) - \liminf_{t\to x} f(t)$.

Proof of the Riemann-Lebesgue Theorem The set D of discontinuity points of $f : [a, b] \to [-M, M]$ naturally filters itself as the countable union

$$D = \bigcup_{k\in\mathbb{N}} D_k$$

where

$$D_k = \{x \in [a, b] : \operatorname{osc}_x(f) \geq \frac{1}{k}\}.$$

According to (a), (c) above, D is a zero set if and only if each D_k is a zero set.

Assume that f is Riemann integrable and let $\epsilon > 0, k \in \mathbb{N}$ be given. By Theorem 19 there is a partition P such that

$$U - L = \sum(M_i - m_i)\Delta x_i < \frac{\epsilon}{k}.$$

Any partition interval $I_i = [x_{i-1}, x_i]$ that contains a point of D_k in its interior has $M_i - m_i \geq 1/k$. Since $\sum(M_i - m_i)\Delta x_i < \epsilon/k$, there cannot be too many such intervals. (This is the key step in the estimates.) More precisely, if we sum over the i's such that I_i contains a point of D_k in its interior then

$$\frac{1}{k}\sum \Delta x_i < \frac{\epsilon}{k}.$$

Except for the zero set of points which lie at partition points, D_k is contained in finitely many open intervals whose total length is $< \epsilon$. Since ϵ is arbitrary, D_k is a zero set. By (c), D is a zero set.

Conversely, assume that the discontinuity set D of $f : [a, b] \to [-M, M]$ is a zero set. Let $\epsilon > 0$ be given. By Riemann's integrability criterion, to prove that f is Riemann integrable it suffices to find $L = L(f, P)$ and $U = U(f, P)$ such that $U - L < \epsilon$. Choose $k \in \mathbb{N}$ so that

$$\frac{1}{k} < \frac{\epsilon}{2(b-a)}.$$

By (a), D_k is a zero set, so there is a countable covering of D_k by open intervals $J_j = (a_j, b_j)$ with total length $\leq \epsilon/4M$. Also, for each $x \in [a, b] \setminus D_k$ there is an open interval I_x containing x such that

$$\sup\{f(t) : t \in I_x\} - \inf\{f(t) : t \in I_x\} < \frac{1}{k}.$$

Consider the collection $\mathcal{V}$ of open intervals J_j and I_x such that $j \in \mathbb{N}$ and $x \in [a, b] \setminus D_k$. It is an open covering of $[a, b]$. Compactness of $[a, b]$ implies that $\mathcal{V}$ has a Lebesgue number $\lambda > 0$.

Let P be any partition of $[a, b]$ having mesh $P < \lambda$. We claim that $U(f, P) - L(f, P) < \epsilon$. Each partition interval I_i is contained wholly in some I_x or wholly in some J_j. (This is what Lebesgue numbers are good for.) Set

$$\mathcal{J} = \{i : I_i \text{ is contained in some } J_j\}.$$

See Figure 73. For some finite m, $J_1 \cup \cdots \cup J_m$ contains those partition

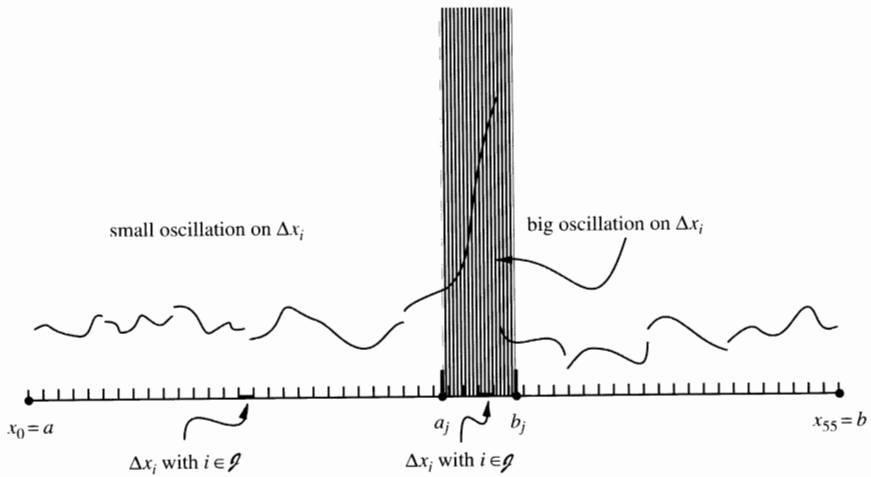


Figure 73 The partition intervals I_i with large oscillation have $i \in \mathcal{J}$.

intervals I_i with $i \in \mathcal{J}$. Also, $\{1, \ldots, n\} = \mathcal{I} \cup \mathcal{J}$. Then

$$\begin{aligned} U - L &= \sum_{i=1}^{n}(M_i - m_i)\Delta x_i \\ &\leq \sum_{i\in\mathcal{J}}(M_i - m_i)\Delta x_i + \sum_{i\notin\mathcal{J}}(M_i - m_i)\Delta x_i \\ &\leq \sum_{i\in\mathcal{J}} 2M\Delta x_i + \sum_{i\notin\mathcal{J}} \frac{1}{k}\Delta x_i \\ &\leq 2M\sum_{j=1}^{m} b_j - a_j + \frac{b-a}{k} \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

For the total length of the intervals I_i contained in the intervals J_j is no greater than $\sum b_j - a_j$. As remarked at the outset, Riemann's integrability criterion then implies that f is integrable. □

The Riemann-Lebesgue Theorem has many consequences, ten of which we list as corollaries.

22 Corollary *Every continuous function is Riemann integrable.*

Proof The discontinuity set of a continuous function is empty, and therefore is a zero set. A continuous function defined on a compact interval $[a, b]$ is bounded. By the Riemann-Lebesgue Theorem, any such function is Riemann integrable. □

23 Corollary *Every piecewise continuous bounded function is Riemann integrable.*

Proof The discontinuity set of a piecewise continuous function is finite and therefore a zero set. By the Riemann-Lebesgue Theorem, such a function is Riemann integrable. □

24 Corollary *Every monotone function is Riemann integrable.*

Proof The set of discontinuities of a monotone function $f : [a, b] \to \mathbb{R}$ is countable and therefore is a zero set. (See Exercise 30 in Chapter 2.) Since f is monotone, its values lie in the interval between $f(a)$ and $f(b)$, so f is bounded. By the Riemann-Lebesgue Theorem, f is Riemann integrable. □

25 Corollary *The product of Riemann integrable functions is Riemann integrable.*

Proof Let $f, g \in \mathcal{R}$ be given. They are bounded and their product is bounded. By the Riemann-Lebesgue Theorem their discontinuity sets, $D(f)$ and $D(g)$, are zero sets, and $D(f) \cup D(g)$ contains the discontinuity set of the product $f \cdot g$. Since the union of two zero sets is a zero set, the Riemann-Lebesgue Theorem implies that $f \cdot g$ is Riemann integrable. □

26 Corollary *If $f : [a, b] \to [c, d]$ is Riemann integrable and $\phi : [c, d] \to \mathbb{R}$ is continuous, then the composite $\phi \circ f$ is Riemann integrable.*

Proof The discontinuity set of $\phi \circ f$ is contained in the discontinuity set of f, and therefore is a zero set. Since ϕ is continuous and $[c, d]$ is compact, $\phi \circ f$ is bounded. By the Riemann-Lebesgue Theorem, $\phi \circ f$ is Riemann integrable. □

27 Corollary *If $f \in \mathcal{R}$ then $|f| \in \mathcal{R}$.*

Proof The function $\phi : y \mapsto |y|$ is continuous, so $x \mapsto |f(x)|$ is Riemann integrable according to Corollary 26. □

28 Corollary *If $a < c < b$ and $f : [a, b] \to \mathbb{R}$ is Riemann integrable then its restrictions to $[a, c]$, $[c, b]$ are Riemann integrable, and*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Conversely, Riemann integrability on $[a, c]$ and $[c, b]$ implies Riemann integrability on $[a, b]$.

Proof See Figure 74. The union of the discontinuity sets for the restrictions of f to the subintervals $[a, c]$, $[c, b]$ is the discontinuity set of f. The latter is a zero set if and only if the former two are, and so by the Riemann-Lebesgue Theorem, f is Riemann integrable if and only if its restrictions to $[a, c]$, $[c, b]$ are.

Let $\chi_{[a,c]}$, $\chi_{[c,b]}$ be the characteristic functions of $[a, c]$, $[c, b]$. By Corollary 23 they are integrable, and by Corollary 25, so are the products $\chi_{[a,c]} \cdot f$ and $\chi_{[c,b]} \cdot f$. Since

$$f = \chi_{[a,c]} \cdot f + \chi_{(c,b]} \cdot f,$$

the addition formula follows from linearity of the integral, Theorem 16. □

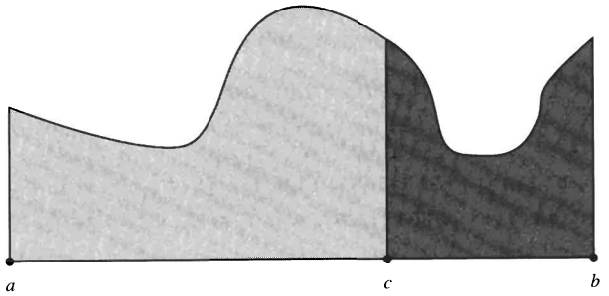


Figure 74 Additivity of the integral is equivalent to additivity of area.

29 Corollary *If $f : [a, b] \to [0, M]$ is Riemann integrable and has integral zero then $f(x) = 0$ at every continuity point x of f. That is, $f(x) = 0$ almost everywhere.*

Proof Suppose not: let x_0 be a continuity point of f and assume that $f(x_0) > 0$. Then for some $\delta > 0$ and each $x \in (x_0 - \delta, x_0 + \delta)$, $f(x) \geq f(x_0)/2$. The function

$$g(x) = \begin{cases} \dfrac{f(x_0)}{2} & \text{if } x \in (x_0 - \delta,\ x_0 + \delta) \\ 0 & \text{otherwise} \end{cases}$$

satisfies $0 \le g(x) \le f(x)$ everywhere. See Figure 75. By monotonicity of the integral, Theorem 17,

$$f(x_0)\delta = \int_a^b g(x)\,dx \le \int_a^b f(x)\,dx = 0,$$

a contradiction. Hence $f(x) = 0$ at every continuity point. □

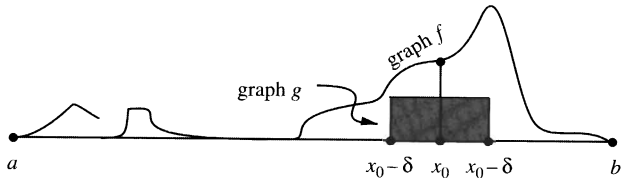


Figure 75 The shaded rectangle prevents the integral of f being zero.

Corollary 26 and Exercises 34, 36, 46, 48 deal with the way that Riemann integrability behaves under composition. If $f \in \mathcal{R}$ and ϕ is continuous then $\phi \circ f \in \mathcal{R}$, although the composition in the other order, $f \circ \phi$, may fail to be integrable. Continuity is too weak a hypothesis for such a "change of variable." See Exercise 36. However, we have the following result.

30 Corollary *If f is Riemann integrable and ψ is a bijection whose inverse satisfies a Lipschitz condition then $f \circ \psi$ is Riemann integrable.*

Proof More precisely, we assume that $f : [a, b] \to \mathbb{R}$ is Riemann integrable, ψ bijects $[c, d]$ onto $[a, b]$, $\psi(c) = a$, $\psi(d) = b$, and for some constant K and all $s, t \in [a, b]$,

$$\left|\psi^{-1}(s) - \psi^{-1}(t)\right| \le K\,|s - t|\,.$$

We then assert that $f \circ \psi$ is a Riemann integrable function $[c, d] \to \mathbb{R}$. Note that ψ^{-1} is a homeomorphism. For it is a continuous bijection whose domain of definition is compact.

Let D be the set of discontinuity points of f. Then $D' = \psi^{-1}(D)$ is the set of discontinuity points for $f \circ \psi$. Let $\epsilon > 0$ be given. There is an open covering of D by intervals (a_i, b_i) whose total length is $\le \epsilon/K$. The homeomorphic intervals $(a_i', b_i') = \psi^{-1}(a_i, b_i)$ cover D' and have total length

$$\sum b_i' - a_i' \le \sum K(b_i - a_i) \le \epsilon.$$

Therefore D' is a zero set and by the Riemann Lebesgue Theorem, $f \circ \psi$ is integrable. □

31 Corollary *If $f \in \mathcal{R}$ and $\psi : [c, d] \to [a, b]$ is a C^1 diffeomorphism then $f \circ \psi$ is Riemann integrable.*

Proof The hypothesis that ψ is a C^1 diffeomorphism means that it is a continuously differentiable bijection, whose derivative is nowhere zero. We assume that $\psi'(t) > 0$. Since ψ' is continuous and positive on $[c, d]$ there is a constant $\kappa > 0$ such that for all $\theta \in [c, d]$, $\psi'(\theta) \ge \kappa$. The Mean Value Theorem implies that for all $u, v \in [c, d]$, there exists a θ between u and v such that

$$\psi(u) - \psi(v) = \psi'(\theta)(u - v).$$

Thus,

$$|\psi(u) - \psi(v)| \ge \kappa\,|u - v|\,.$$

For any $s, t \in [a, b]$, set $u = \psi^{-1}(s)$, $v = \psi^{-1}(t)$. Then

$$|s - t| \ge \kappa\left|\psi^{-1}(s) - \psi^{-1}(t)\right|,$$

which is a Lipschitz condition on ψ^{-1} with Lipschitz constant $K = \kappa^{-1}$. By Corollary 30, $f \circ \psi$ is Riemann integrable. □

Versions of the preceding theorem and corollary remain true without the hypotheses that ψ bijects. The proofs are harder because ψ can fold infinitely often. See Exercise 39.

In calculus you learn that the derivative of the integral is the integrand. This we now prove.

32 Fundamental Theorem of Calculus *If $f : [a, b] \to \mathbb{R}$ is Riemann integrable then its indefinite integral*

$$F(x) = \int_a^x f(t)\,dt$$

is a continuous function of x. The derivative of $F(x)$ exists and equals $f(x)$ at all points x at which f is continuous.

Proof #1 Obvious from Figure 76.

Proof #2 Since f is Riemann integrable, it is bounded; say $|f(x)| \le M$ for all x. By Corollary 28

$$|F(y) - F(x)| = \left|\int_x^y f(t)\,dt\right| \le M\,|y - x|\,.$$

Therefore, F is continuous: given $\epsilon > 0$, choose $\delta < \epsilon/M$, and observe that $|y - x| < \delta$ implies that $|F(y) - F(x)| < M\delta < \epsilon$. In exactly the same way, if f is continuous at x then

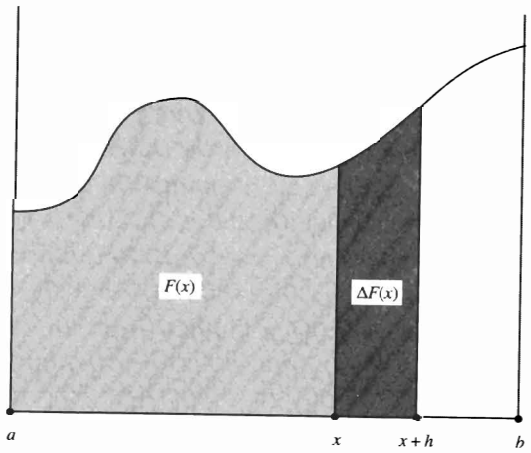


Figure 76 Why *does* this picture give a proof of the Fundamental Theorem of Calculus?

$$\frac{F(x+h)-F(x)}{h} = \frac{1}{h}\int_x^{x+h} f(t)\,dt \to f(x)$$

as $h \to 0$. For if

$$m(x,h) = \inf\{f(s) : |s-x| \le |h|\}$$
$$M(x,h) = \sup\{f(s) : |s-x| \le |h|\}$$

then

$$m(x,h) = \frac{1}{h}\int_x^{x+h} m(x,h)\,dt \le \frac{1}{h}\int_x^{x+h} f(t)\,dt$$
$$\le \frac{1}{h}\int_x^{x+h} M(x,h)\,dt = M(x,h).$$

When f is continuous at x, $m(x,h)$ and $M(x,h)$ converge to $f(x)$ as $h \to 0$, and so must the integral sandwiched between them,

$$\frac{1}{h}\int_x^{x+h} f(t)\,dt.$$

(If $h < 0$ then $\frac{1}{h}\int_x^{x+h} f(t)\,dt$ is interpreted as $-\frac{1}{h}\int_{x+h}^{x} f(t)\,dt$.) □

33 Corollary *The derivative of an indefinite Riemann integral exists almost everywhere and equals the integrand almost everywhere.*

Proof Assume that $f : [a,b] \to \mathbb{R}$ is Riemann integrable and $F(x)$ is its indefinite integral. By the Riemann-Lebesgue Theorem, f is continuous almost everywhere, and by the Fundamental Theorem of Calculus, $F'(x)$ exists and equals $f(x)$ wherever f is continuous. □

A second version of the Fundamental Theorem of Calculus concerns antiderivatives. If one function is the derivative of another, the second function is an **antiderivative** of the first.

Note When G is an antiderivative of of $g : [a,b] \to \mathbb{R}$, we have

$$G'(x) = g(x)$$

for *all* $x \in [a,b]$, not merely for almost all $x \in [a,b]$.

34 Corollary *Every continuous function has an antiderivative.*

Proof Assume that $f : [a,b] \to \mathbb{R}$ is continuous. By the Fundamental Theorem of Calculus, the indefinite integral $F(x)$ has a derivative everywhere, and $F'(x) = f(x)$ everywhere. □

Some discontinuous functions have an antiderivative and others don't. Surprisingly, the wildly oscillating function

$$f(x) = \begin{cases} 0 & \text{if } x \le 0 \\ \sin\dfrac{\pi}{x} & \text{if } x > 0 \end{cases}$$

has an antiderivative, but the jump function

$$g(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 1 & \text{if } x > 0 \end{cases}$$

does not. See Exercise 42.

35 Antiderivative Theorem *An antiderivative of a Riemann integrable function, if it exists, differs from the indefinite integral by a constant.*

Proof We assume that $f : [a,b] \to \mathbb{R}$ is Riemann integrable, that G is an antiderivative of f, and we assert that for all $x \in [a,b]$,

$$G(x) = \int_a^x f(t)\,dt + C,$$

where C is a constant. (In fact, $C = G(a)$.) Partition $[a, x]$ as

$$a = x_0 < x_1 < \cdots < x_n = x,$$

and choose $t_k \in [x_{k-1}, x_k]$ such that

$$G(x_k) - G(x_{k-1}) = G'(t_k)\Delta x_k.$$

Such a t_k exists by the Mean Value Theorem applied to the differentiable function G. Telescoping gives

$$G(x) - G(a) = \sum_{k=1}^{n} G(x_k) - G(x_{k-1}) = \sum_{k=1}^{n} f(t_k)\Delta x_k,$$

which is a Riemann sum for f on the interval $[a, x]$. Since f is Riemann integrable, the Riemann sum converges to $F(x)$ as the mesh of the partition tends to zero. This gives $G(x) - G(a) = F(x)$ as claimed. □

36 Corollary *Standard integral formulas, such as*

$$\int_a^b x^2\,dx = \frac{b^3 - a^3}{3},$$

are valid.

Proof Every integral formula is actually a derivative formula, and the Anti-derivative Theorem converts derivative formulas to integral formulas. □

In particular, the **logarithm function** is defined as the integral,

$$\log x = \int_1^x \frac{1}{t}\,dt.$$

Since the integrand $1/t$ is well defined and continuous when $t > 0$, $\log x$ is well defined and differentiable for $x > 0$. Its derivative is $1/x$. By the way, as is standard in post calculus vocabulary, $\log x$ refers to the natural logarithm, not to the base 10 logarithm. See also Exercise 16.

An antiderivative of f has $G'(x) = f(x)$ everywhere, and differs from the indefinite integral $F(x)$ by a constant. But what if we assume instead that $H'(x) = f(x)$ *almost* everywhere? Should this not also imply $H(x)$ differs from $F(x)$ by a constant? Surprisingly, the answer is "no."

37 Theorem *There exists a continuous function $H : [0, 1] \to \mathbb{R}$ whose derivative exists and equals zero almost everywhere, but which is not constant.*

Proof. The counter-example is the **Devil's staircase function**, also called the **Cantor function**. It is defined as

$$H(x) = \begin{cases} 1/2 & \text{if } x \in [1/3, 2/3] \\ 1/4 & \text{if } x \in [1/9, 2/9] \\ 3/4 & \text{if } x \in [7/9, 8/9] \\ \cdots & \cdots \end{cases}$$

See Figures 77, 78.

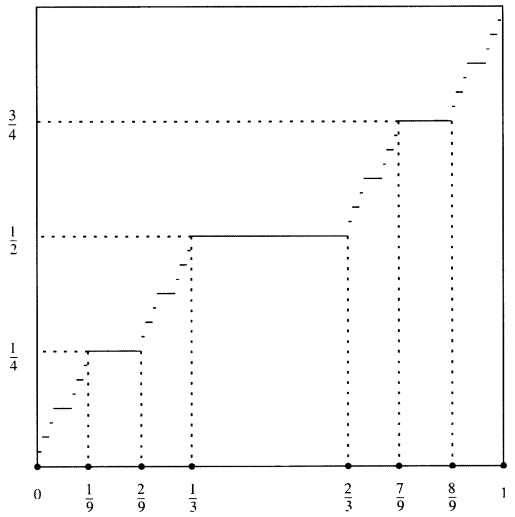


Figure 77 The devil's staircase.

On each discarded interval in the Cantor set construction, H is constant. Thus H is differentiable with derivative zero at all points of $[0, 1] \setminus C$, and since C is a zero set, this implies $H'(x) = 0$ for almost every x.

To show that H is continuous we use base 2 and base 3 arithmetic. If $x \in [0, 1]$ has base 3 expansion $x = (.x_1x_2\ldots)_3$ then the base 2 expansion of $y = H(x)$ is $(.y_1y_2\ldots)_2$ where

$$y_i = \begin{cases} 0 & \text{if } \exists k < i \text{ such that } x_k = 1 \\ 1 & \text{if } x_i = 1 \text{ and } \nexists k < i \text{ such that } x_k = 1 \\ \dfrac{x_i}{2} & \text{if } x_i = 0 \text{ or } x_i = 2 \text{ and } \nexists k < i \text{ such that } x_k = 1. \end{cases}$$

You should ask: is this a valid definition of $H(x)$? If two different base 3 expansions, $(.x_1x_2\ldots)_3$ and $(.x_1'x_2'\ldots)_3$ represent the same number x, do the two base 2 expressions for $H(x)$ represent the same number y? Two

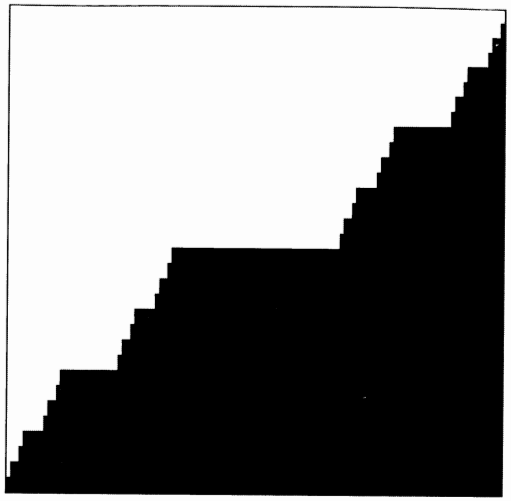


Figure 78 The devil's undergraph.

base 3 expansions of x represent the same number x if and only if x is an endpoint of C; one of its base 3 expansions ends in 0's, the other in 2's. For example,

$$(.x_1x_2\ldots x_\ell 0\overline{2})_3 = x = (.x_1x_2\ldots x_\ell 1\overline{0})_3.$$

If for some (smallest) $k \le \ell$, $x_k = 1$ then

$$(.y_1y_2\ldots)_2 = \left(.\frac{x_1}{2}\,\frac{x_2}{2}\,\ldots\,\frac{x_{k-1}}{2}\,1\overline{0}\right)_2$$

unambiguously. If none of x_k with $k \le \ell$ equals 1 then the two base 2 expansions corresponding to $H(x)$ are

$$\left(.\frac{x_1}{2}\,\frac{x_2}{2}\,\ldots\,\frac{x_{\ell-1}}{2}\,0\overline{1}\right)_2$$
$$\left(.\frac{x_1}{2}\,\frac{x_2}{2}\,\ldots\,\frac{x_{\ell-1}}{2}\,1\overline{0}\right)_2.$$

These two base 2 expansions represent the same number y. The same reasoning applies to the only other ambiguous case,

$$(.x_1x_2\ldots x_\ell 1\overline{2})_3 = x = (.x_1x_2\ldots x_\ell 2\overline{0})_3.$$

The point $y = H(x)$ is well defined.

Continuity is now easy to check. Let $\epsilon > 0$ be given. Choose k such that $1/2^k \le \epsilon$. If $|x - x'| < 1/3^k$ then there are base 3 expansions of x, x' whose first k symbols agree. Therefore the first k symbols of $H(x)$ and $H(x')$ agree, which implies that

$$|H(x) - H(x')| \le \frac{1}{2^{k+1}} < \epsilon. \qquad \square$$

A yet more pathological example is a *strictly* monotone, continuous function J whose derivative is almost everywhere zero. Its graph is a sort of **Devil's ski slope**, almost everywhere level but also everywhere downhill. To construct J, start with H and extend it to a function $\widehat{H} : \mathbb{R} \to \mathbb{R}$ by setting $\widehat{H}(x + n) = H(x) + n$ for all $n \in \mathbb{Z}$ and all $x \in [0, 1]$. Then set

$$J(x) = \sum_{k=0}^{\infty} \frac{\widehat{H}(3^k x)}{4^k}.$$

The values of $\widehat{H}(3^k x)$ for $x \in [0, 1]$ are $\le 3^k$, which is much smaller than the denominator 4^k. Thus the series converges and $J(x)$ is well defined. According to the Weierstrass M-test, proved in the next chapter, J is continuous. Since $\widehat{H}(3^k x)$ strictly increases for any pair of points at distance $> 1/3^k$ apart, and this fact is preserved when we take sums, J strictly increases. The proof that $J'(x) = 0$ almost everywhere requires more and deeper theory.

Next, we justify two common methods of integration.

38 Integration by substitution *If $f \in \mathcal{R}$ and $g : [c, d] \to [a, b]$ is a continuously differentiable bijection with $g' > 0$ (g is a C^1 diffeomorphism) then*

$$\int_a^b f(y)\,dy = \int_c^d f(g(x))g'(x)\,dx.$$

Proof The first integral exists by assumption. By Corollary 31, the composite $f \circ g \in \mathcal{R}$, and since g' is continuous, the second integral exists by Corollary 25. To show that the two integrals are equal we resort again to Riemann sums. Let P partition the interval $[c, d]$ as

$$c = x_0 < x_1 < \cdots < x_n = d$$

and choose $t_k \in [x_{k-1}, x_k]$ such that

$$g(x_k) - g(x_{k-1}) = g'(t_k)\Delta x_k.$$

The Mean Value Theorem ensures that such a t_k exists. Since g is a diffeomorphism we have a partition Q of the interval $[a, b]$

$$a = y_0 < y_1 < \cdots < y_n = b$$

where $y_k = g(x_k)$, and $\|P\| \to 0$ implies that $\|Q\| \to 0$. Set $s_k = g(t_k)$. This gives two equal Riemann sums

$$\sum_{k=1}^{n} f(s_k)\,\Delta y_k = \sum_{k=1}^{n} f(g(t_k))g'(t_k)\,\Delta x_k$$

which converge to the integrals $\int_a^b f(y)\,dy$ and $\int_c^d f(g(t)g'(t)\,dt$ as $\|P\| \to 0$. Since the limits of equals are equal, the integrals are equal. □

Actually, it is sufficient to assume that $g' \in \mathcal{R}$.

39 Integration by Parts *If $f, g : [a, b] \to \mathbb{R}$ are differentiable and $f', g' \in \mathcal{R}$ then*

$$\int_a^b f(x)g'(x)\,dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\,dx.$$

Proof Differentiability implies continuity implies integrability, so $f, g \in \mathcal{R}$. Since the product of Riemann integrable functions is Riemann integrable, $f'g, fg' \in \mathcal{R}$, and both integrals exist. By the Leibniz Rule, $(fg)'(x) = f(x)g'(x) + f'(x)g(x)$ everywhere. That is, fg is an antiderivative of $f'g + fg'$. The Antiderivative Theorem states that fg differs from the indefinite integral of $f'g + fg'$ by a constant. That is, for all $t \in [a, b]$,

$$\begin{aligned} f(t)g(t) - f(a)g(a) &= \int_a^t f'(x)g(x) + f(x)g'(x)\,dx \\ &= \int_a^t f'(x)g(x)\,dx + \int_a^t f(x)g'(x)\,dx. \end{aligned}$$

Setting $t = b$ gives the result. □

Improper Integrals

Assume that $f : [a, b) \to \mathbb{R}$ is Riemann integrable when restricted to any closed subinterval $[a, c] \subset [a, b)$. You may imagine that $f(x)$ has some unpleasant behavior as $x \to b$, such as $\limsup_{x\to b} |f(x)| = \infty$ and/or $b = \infty$. See Figure 79.

If the limit of $\int_a^c f(x)\,dx$ exists (and is a real number) as $c \to b$ then it is natural to define it as the **improper Riemann integral**

$$\int_a^b f(x)\,dx = \lim_{c\to b} \int_a^c f(x)\,dx.$$

The same idea works of course on an interval to the left of a. In order that the two sided improper integral exists for a function $f : (a, b) \to \mathbb{R}$ it is

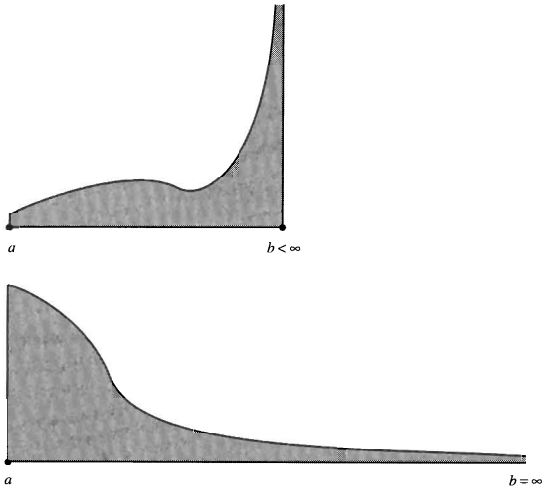


Figure 79 The improper integral converges if and only if the total undergraph area is finite.

natural to fix some point $m \in (a, b)$ and require that both improper integrals $\int_a^m f(x)\,dx$ and $\int_m^b f(x)\,dx$ exist. Their sum is the improper integral $\int_a^b f(x)\,dx$. With some ingenuity you can devise a function $f : \mathbb{R} \to \mathbb{R}$ whose improper integral $\int_{-\infty}^{\infty} f(x)\,dx$ exists despite the fact that it is unbounded at both $\pm\infty$. See Exercise 71.

3 Series

A series is a formal sum $\sum a_k$ where the terms a_k are real numbers. The n^{th} partial sum of the series is

$$A_n = a_0 + a_1 + a_2 + \cdots + a_n.$$

The series **converges** to A if $A_n \to A$ as $n \to \infty$, and we write

$$A = \sum_{k=0}^{\infty} a_k.$$

A series that does not converge **diverges**. The basic question to ask about a series is: does it converge or diverge?

For example, if λ is a constant and $|\lambda| < 1$ then the **geometric series**

$$\sum_{k=0}^{\infty} \lambda^k = 1 + \lambda + \cdots + \lambda^n + \ldots$$

converges to $1/(1-\lambda)$. For its partial sums are

$$\Lambda_n = 1 + \lambda + \lambda^2 + \cdots + \lambda^n = \frac{1-\lambda^{n+1}}{1-\lambda}$$

and $\lambda^{n+1} \to 0$ as $n \to \infty$. On the other hand, if $|\lambda| \geq 1$, the series $\sum \lambda^k$ diverges.

Let $\sum a_n$ be a series. The Cauchy Convergence Criterion from Chapter 1 applied to its sequence of partial sums yields the **CCC for series**

$\sum a_k$ converges if and only if

$$\forall \epsilon > 0 \ \ \exists N \text{ such that } m, n \geq N \quad \Rightarrow \quad \left| \sum_{k=m}^{n} a_k \right| < \epsilon.$$

One immediate consequence of the CCC is that no finite number of terms affects convergence of a series. Rather, it is the **tail** of the series, the terms a_k with k large, that determines convergence or divergence. Likewise, whether the series leads off with a term of index $k = 0$ or $k = 1$, etc. is irrelevant.

A second consequence of the CCC is that if a_k does not converge to zero as $k \to \infty$ then $\sum a_k$ does not converge. For Cauchyness of the partial sum sequence (A_n) implies that $a_n = A_n - A_{n-1}$ becomes small when $n \to \infty$. If $|\lambda| \geq 1$ the geometric series $\sum \lambda^k$ diverges since its terms do not converge to zero. The **harmonic series**

$$\sum_{k=1}^{\infty} \frac{1}{k} = 1 + \frac{1}{2} + \frac{1}{3} + \ldots$$

gives an example that a series can diverge even though its terms do tend to zero. See below.

Series theory has a large number of convergence tests. All boil down to the following result.

40 Comparison Test *If a series $\sum b_k$ **dominates** a series $\sum a_k$ in the sense that for all sufficiently large k, $|a_k| \leq b_k$, then convergence of $\sum b_k$ implies convergence of $\sum a_k$.*

Proof Given $\epsilon > 0$, convergence of $\sum b_k$ implies there is a large N such that for all $m, n \geq N$, $\sum_{k=m}^{n} b_k < \epsilon$. Thus

$$\left| \sum_{k=m}^{n} a_k \right| \leq \sum_{k=m}^{n} |a_k| \leq \sum_{k=m}^{n} b_k < \epsilon,$$

and convergence of $\sum a_k$ follows from the CCC. □

Example The series $\sum \sin(k)/2^k$ converges since it is dominated by the geometric series $\sum 1/2^k$.

A series $\sum a_k$ converges **absolutely** if $\sum |a_k|$ converges. The comparison test shows that absolute convergence implies convergence. A series that converges but not absolutely converges **conditionally**: $\sum a_k$ converges and $\sum |a_k|$ diverges. See below.

Series and integrals are both infinite sums. You can imagine a series as an improper integral in which the integration variable is an integer,

$$\sum_{k=0}^{\infty} a_k = \int_{\mathbb{N}} a_k \, dk.$$

More precisely, given a series $\sum a_k$, define $f : [0, \infty) \to \mathbb{R}$ by setting

$$f(x) = a_k \text{ if } k - 1 < x \leq k.$$

See Figure 80. Then

$$\sum_{k=0}^{\infty} a_k = \int_0^{\infty} f(x) \, dx.$$

The series converges if and only if the improper integral does. The natural extension of this picture is the

41 Integral Test *Suppose that $\int_0^{\infty} f(x)\,dx$ is a given improper integral and and $\sum a_k$ is a given series.*

(a) If $|a_k| \leq f(x)$ for all sufficiently large k and all $x \in (k-1, k]$ then convergence of the improper integral implies convergence of the series.

(b) If $|f(x)| \leq a_k$ for all sufficiently large k and all $x \in [k, k+1)$ then divergence of the improper integral implies divergence of the series.

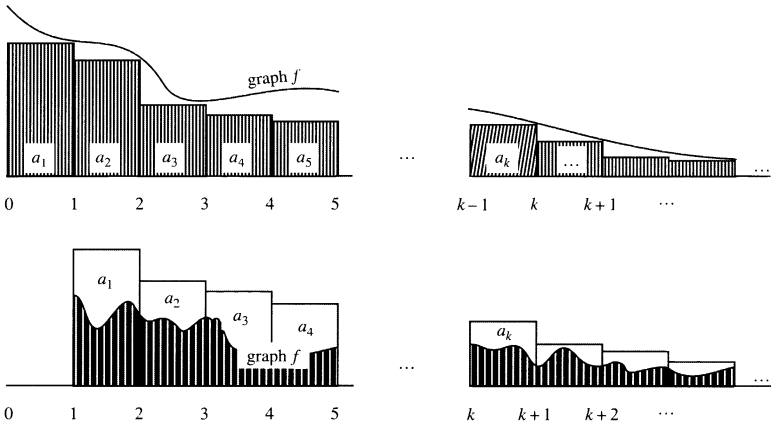


Figure 80 The pictorial proof of the integral test.

Proof See Figure 80.

(a) For some large N_0 and all $N \geq N_0$ we have

$$\sum_{k=N_0+1}^{N} |a_k| \leq \int_{N_0}^{N} f(x)\,dx \leq \int_{0}^{\infty} f(x)\,dx,$$

which is a finite real number. An increasing, bounded sequence converges to a limit, so the tail of the series $\sum |a_k|$ converges, and the whole series $\sum |a_k|$ converges. Absolute convergence implies convergence.

The proof of (b) is left as Exercise 73. □

Example The ***p*-series**, $\sum 1/k^p$ converges when $p > 1$ and diverges when $p \leq 1$.

Case 1. $p > 1$. By the fundamental theorem of calculus and differentiation rules

$$\int_1^b \frac{1}{x^p} = \frac{b^{1-p} - 1}{1-p} \to \frac{1}{p-1}$$

as $b \to \infty$. The improper integral converges and dominates the p-series, which implies convergence of the series by the integral test.

Case 2. $p \leq 1$. The p-series dominates the improper integral

$$\int_1^b \frac{1}{x^p}\,dx = \begin{cases} \log b & \text{if } p = 1 \\ \dfrac{b^{1-p} - 1}{1-p} & \text{if } p < 1. \end{cases}$$

As $b \to \infty$, these quantities blow up, and the integral test implies divergence of the series. When $p = 1$ we have the harmonic series, which we have just shown to diverge.

The **exponential growth rate** of the series $\sum a_k$ is

$$\alpha = \limsup_{k\to\infty} \sqrt[k]{|a_k|}.$$

42 Root Test *Let α be the exponential growth rate of a series $\sum a_k$. If $\alpha < 1$ the series converges, if $\alpha > 1$ the series diverges, and if $\alpha = 1$ the root test is inconclusive.*

Proof If $\alpha < 1$, fix a constant β,

$$\alpha < \beta < 1.$$

Then for all large k, $|a_k|^{1/k} \leq \beta$; i.e.,

$$|a_k| \leq \beta^k,$$

which gives convergence of $\sum a_k$ by comparison to the geometric series $\sum \beta^k$.

If $\alpha > 1$, choose β, $1 < \beta < \alpha$. Then $|a_k| \geq \beta^k$ for infinitely many k. Since the terms a_k do not converge to 0, the series diverges.

To show that the root test is inconclusive when $\alpha = 1$, we must find two series, one convergent and the other divergent, both having exponential growth rate $\alpha = 1$. The examples are p-series. We have

$$\log\left(\frac{1}{k^p}\right)^{1/k} = \frac{-p\log(k)}{k} \sim \frac{-p\log(x)}{x} \sim \frac{-p/x}{1} \sim 0$$

by L'Hospital's rule as $k = x \to \infty$. Therefore $\alpha = \lim_{k\to\infty}(1/k^p)^{1/k} = 1$. Since the square series $\sum 1/k^2$ converges and the harmonic series $\sum 1/k$ diverges the root test is inconclusive when $\alpha = 1$. □

43 Ratio Test *Let the ratio between successive terms of the series $\sum a_k$ be $r_k = |a_{k+1}/a_k|$, and set*

$$\liminf_{k\to\infty} r_k = \lambda \qquad \limsup_{k\to\infty} r_k = \rho.$$

If $\rho < 1$ the series converges, if $\lambda > 1$ the series diverges, and otherwise the ratio test is inconclusive.

Proof If $\rho < 1$, choose β, $\rho < \beta < 1$. For all $k \geq$ some K, $|a_{k+1}/a_k| < \beta$; i.e.,

$$|a_k| \leq \beta^{k-K}|a_K| = C\beta^k$$

where $C = \beta^{-K}|a_K|$ is a constant. Convergence of $\sum a_k$ follows from comparison with the geometric series $\sum C\beta^k$. If $\lambda > 1$, choose β, $1 < \beta < \lambda$. Then $|a_k| \geq \beta^k/C$ for all large k, and $\sum a_k$ diverges because its terms do not converge to 0. Again the p-series all have ratio limit $\rho = \lambda = 1$ and demonstrate the inconclusiveness of the ratio test when $\rho = 1$ or $\lambda = 1$. □

Although it is usually easier to apply the ratio test than the root test, the latter has a strictly wider scope. See Exercises 56, 60.

Conditional Convergence

If (a_k) is a decreasing sequence in $\mathbb{R}$ that converges to 0 then its **alternating series**

$$\sum(-1)^{k+1}a_k = a_1 - a_2 + a_3 - \dots$$

converges. For,

$$A_{2n} = (a_1 - a_2) + (a_3 - a_4) + \dots (a_{2n-1} - a_{2n}).$$

and $a_{k-1} - a_k$ is the length of the the interval $I_k = (a_k, a_{k-1})$. The intervals I_k are disjoint, so the sum of their lengths is at most the length of $(0, a_0)$, namely a_0. See Figure 81.

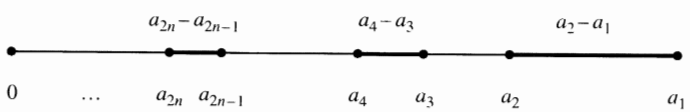


Figure 81 The pictorial proof of alternating convergence.

The sequence (A_{2n}) is increasing and bounded, so $\lim_{n\to\infty} A_{2n}$ exists. The partial sum A_{2n+1} differs from A_{2n} by a_{2n+1}, a quantity that converges to 0 as $n \to \infty$, so

$$\lim_{n\to\infty} A_{2n} = \lim_{n\to\infty} A_{2n+1},$$

and the alternating series converges.

When $a_k = 1/k$ we have the **alternating harmonic series**,

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$$

which we have just shown is convergent.

Series of Functions

A series of functions is of the form

$$\sum_{k=0}^{\infty} f_k(x),$$

where each term $f_k : (a, b) \to \mathbb{R}$ is a function. For example, in a power series

$$\sum c_k x^k$$

the functions are monomials $c_k x^k$. (The coefficients c_k are constants and x is a real variable.) If you think of $\lambda = x$ as a variable, then the geometric series is a power series whose coefficients are 1, $\sum x^k$. Another example of a series of functions is a Fourier series

$$\sum a_k \sin(kx) + b_k \cos(kx).$$

44 Radius of Convergence Theorem *If $\sum c_k x^k$ is a power series then there is a unique R, $0 \leq R \leq \infty$, its **radius of convergence**, such that the series converges whenever $|x| < R$, and diverges whenever $|x| > R$. Moreover R is given by the formula*

$$R = \frac{1}{\limsup_{k\to\infty} \sqrt[k]{|c_k|}}.$$

Proof Apply the root test to the series $\sum c_k x^k$. Then

$$\limsup_{k\to\infty} \sqrt[k]{|c_k x^k|} = |x| \limsup_{k\to\infty} \sqrt[k]{|c_k|} = \frac{|x|}{R}.$$

If $|x| < R$ the root test gives convergence. If $|x| > R$ it gives divergence. □

There are power series with any given radius of convergence, $0 \leq R \leq \infty$. The series $\sum k^k x^k$ has $R = 0$. The series $\sum x^k/\sigma^k$ has $R = \sigma$ for $0 < \sigma < \infty$. The series $\sum x^k/k!$ has $R = \infty$. Eventually, we show that a function defined by a power series is analytic: it is has all derivatives at all points and it can be expanded as a Taylor series at each point inside its radius of convergence, not merely at x = 0. See Section 6 in Chapter 4.

Exercises

1. Assume that $f : \mathbb{R} \to \mathbb{R}$ satisfies $|f(t) - f(x)| \leq |t - x|^2$ for all t, x. Prove that f is constant.
2. A function $f : (a, b) \to \mathbb{R}$ satisfies a **Hölder condition of order α** if $\alpha > 0$, and for some constant H and all $u, x \in (a, b)$,
$$|f(u) - f(x)| \leq H\,|u - x|^{\alpha}\,.$$
The function is said to be **α-Hölder**, with α-Hölder constant H. (The terms "Lipschitz function of order α" and "α-Lipschitz function" are sometimes used with the same meaning.)
   (a) Prove that an α-Hölder function defined on (a, b) is uniformly continuous and infer that it extends uniquely to a continuous function defined on $[a, b]$. Is the extended function α-Hölder?
   (b) What does α-Hölder continuity mean when $\alpha = 1$?
   (c) Prove that α-Hölder continuity when $\alpha > 1$ implies that f is constant.
3. Assume that $f : (a, b) \to \mathbb{R}$ is differentiable.
   (a) If $f'(x) > 0$ for all x, prove that f is strictly monotone increasing.
   (b) If $f'(x) \geq 0$ for all x, what can you prove?
4. Prove that $\sqrt{n+1} - \sqrt{n} \to 0$ as $n \to \infty$.
5. Assume that $f : \mathbb{R} \to \mathbb{R}$ is continuous, and for all $x \neq 0$, $f'(x)$ exists. If $\lim_{x \to 0} f'(x) = L$ exists, does it follow that $f'(0)$ exists? Prove or disprove.
6. If a differentiable function $f : (a, b) \to \mathbb{R}$ assumes a maximum or a minimum at some $\theta \in (a, b)$, prove that $f'(\theta) = 0$. Why is the assertion false when $[a, b]$ replaces (a, b)?
7. In L'Hospital's Rule, replace the interval (a, b) with the half-line $[a, \infty)$ and interpret "x tends to b" as "$x \to \infty$." Show that if f/g tends to $0/0$ and f'/g' tends to L then f/g also tends to L also. Prove that this continues to hold when $L = \infty$ in the sense that if $f'/g' \to \infty$ then $f/g \to \infty$.
8. In L'Hospital's Rule, replace the assumption that f/g tends to $0/0$ with the assumption that it tends to ∞/∞. If f'/g' tends to L, prove that f/g tends to L also. [Hint: Think of a rear guard instead of an advance guard.] [Query: Is there a way to deduce the ∞/∞ case from the $0/0$ case? Naïvely taking reciprocals does not work.]
9. (a) Draw the graph of a continuous function defined on $[0, 1]$ that is differentiable on the interval $(0, 1)$ but not at the endpoints.

   (b) Can you find a formula for such a function?
   (c) Does the Mean Value Theorem apply to such a function?
10. Let $f : (a, b) \to \mathbb{R}$ be given.
   (a) If $f''(x)$ exists, prove that
$$\lim_{h \to 0} \frac{f(x-h) - 2f(x) + f(x+h)}{h^2} = f''(x).$$
   (b) Find an example that this limit can exist even when $f''(x)$ fails to exist.
11. Assume that $f : (-1, 1) \to \mathbb{R}$ and $f'(0)$ exists. If $\alpha_n, \beta_n \to 0$ as $n \to \infty$, define the difference quotient
$$D_n = \frac{f(\beta_n) - f(\alpha_n)}{\beta_n - \alpha_n}.$$
   (a) Prove that $\lim_{n \to \infty} D_n = f'(0)$ under each of the following conditions
      (i) $\alpha_n < 0 < \beta_n$.
      (ii) $0 < \alpha_n < \beta_n$ and $\dfrac{\beta_n}{\beta_n - \alpha_n} \leq M$.
      (iii) $f'(x)$ exists and is continuous for all $x \in (-1, 1)$.
   (b) Set $f(x) = x^2 \sin(1/x)$ for $x \neq 0$ and $f(0) = 0$. Observe that f is differentiable everywhere in $(-1, 1)$ and $f'(0) = 0$. Find α_n, β_n that tend to 0 in such a way that D_n converges to a limit unequal to $f'(0)$.
12. Assume that f and g are r^{th} order differentiable functions $(a, b) \to \mathbb{R}$, $r \geq 1$. Prove the r^{th} order Leibniz product rule for the function $f \cdot g$,
$$(f \cdot g)^{(r)}(x) = \sum_{k=0}^{r} \binom{r}{k} f^{(k)}(x) \cdot g^{(r-k)}(x).$$
where $\binom{r}{k} = r!/(k!(r-k)!)$ is the binomial coefficient, r choose k. [Hint: Induction.]
13. Assume that $f : \mathbb{R} \to \mathbb{R}$ is differentiable.
   (a) If there is an $L < 1$ such that for each $x \in \mathbb{R}$, $f'(x) < L$, prove that there exists a unique point x such that $f(x) = x$. [x is a fixed point for f.]
   (b) Show by example that (a) fails if $L = 1$.
14. Define $e : \mathbb{R} \to \mathbb{R}$ by
$$e(x) = \begin{cases} e^{-1/x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

(a) Prove that e is smooth; that is, e has derivatives of all orders at all points x. [Hint: L'Hospital and induction. Feel free to use the standard differentiation formulas about e^x from calculus.]
(b) Is e analytic?
(c) Show that the **bump function**

$$\beta(x) = e^2 e(1-x) \cdot e(x+1)$$

is smooth, identically zero outside the interval $(-1, 1)$, positive inside the interval $(-1, 1)$, and takes value 1 at $x = 0$. (e^2 is the square of the base of the natural logarithms, while $e(x)$ is the function just defined. Apologies to the abused notation.)
(d) For $|x| < 1$ show that

$$\beta(x) = e^{-2x^2/(x^2-1)}.$$

Bump functions have wide use in smooth function theory and differential topology. The graph of β looks like a bump. See Figure 82.

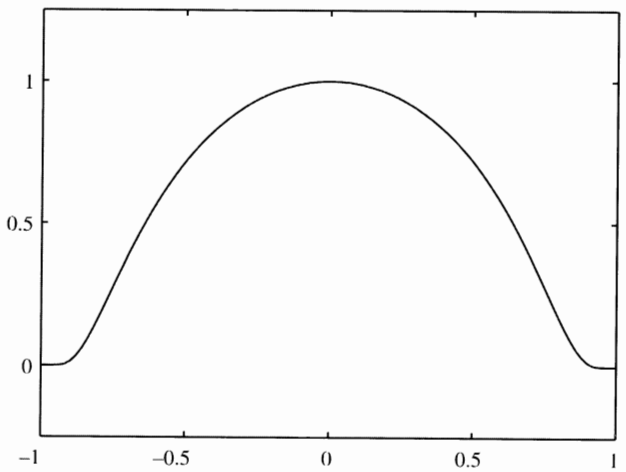


Figure 82 The graph of the bump function β.

**15. Let L be any closed set in $\mathbb{R}$. Prove that there is a smooth function $f : \mathbb{R} \to [0, 1]$ such that $f(x) = 0$ if and only if $x \in L$. To put it another way, every closed set in $\mathbb{R}$ is the zero locus of some smooth function. [Hint: Use Exercise 14(c).]

16. $\log x$ is defined to be $\int_1^x 1/t\, dt$ for $x > 0$. Using only the mathematics explained in this chapter,

(a) Prove that log is a smooth function.
(b) Prove that $\log(xy) = \log x + \log y$ for all $x, y > 0$. [Hint: Fix y and define $f(x) = \log(xy) - \log x - \log y$. Show that $f(x) \equiv 0$.]
(c) Prove that log is strictly monotone increasing and its range is all of $\mathbb{R}$.

17. Define $f(x) = x^2$ if $x < 0$ and $f(x) = x + x^2$ if $x \geq 0$. Differentiation gives $f''(x) \equiv 2$. This is bogus. Why?

18. Recall that the κ-oscillation set of an arbitrary function $f : [a, b] \to \mathbb{R}$ is

$$D_\kappa = \{x \in [a, b] : \operatorname{osc}_x(f) \geq \kappa\}.$$

(a) Prove that D_κ is closed.
(b) Infer that the discontinuity set of f is a countable union of closed sets. (This is called an F_σ-set.)
(c) Infer from (b) that the set of continuity points is a countable intersection of open sets. (This is called a G_δ-set.)

*19. Baire's Theorem (page 243) asserts that if a complete metric space is the countable union of closed subsets then at least one of them has non-empty interior. Use Baire's Theorem to show that the set of irrational numbers is not the countable union of closed subsets of $\mathbb{R}$.

20. Use Exercises 18 and 19 to show that there is no function $f : \mathbb{R} \to \mathbb{R}$ which is discontinuous at every irrational number and continuous at every rational number.

**21. Find a subset S of the middle-thirds Cantor set which is never the discontinuity set of a function $f : \mathbb{R} \to \mathbb{R}$. Infer that some zero sets are never discontinuity sets of Riemann integrable functions. [Hint: How many subsets of C are there? How many can be countable unions of closed sets?]

22. Suppose that $f_n : [a, b] \to \mathbb{R}$ is a sequence of continuous functions that converges pointwise to a limit function $f : [a, b] \to \mathbb{R}$. Such an f is said to be of **Baire class 1. (Pointwise convergence is discussed in the next chapter. It means what it says: for each x, $f_n(x)$ converges to $f(x)$ as $n \to \infty$. Continuous functions are considered to be of Baire class 0, and in general a Baire class k function is the pointwise limit of a sequence of Baire class $k-1$ functions. Strictly speaking, it should not be of Baire class $k-1$ itself, but for simplicity I include continuous functions among Baire class 1 functions. It is an interesting fact that for every k there are Baire class k functions not of Baire class $k-1$. You might consult *A Primer of Real Functions* by Ralph Boas.)

Prove that the κ-oscillation set of f is nowhere dense, as follows. To arrive at a contradiction, suppose that D_κ is dense in some interval $(\alpha, \beta) \subset [a, b]$. By Exercise 18, D_κ is closed, so it contains (α, β). Cover $\mathbb{R}$ by countably many intervals (a_ℓ, b_ℓ) of length $< \kappa$ and set

$$H_\ell = f^{\text{pre}}(a_\ell, b_\ell).$$

(a) Why does $\bigcup_\ell H_\ell = [a, b]$?
(b) Show that no H_ℓ contains a subinterval of (α, β).
(c) Why are

$$F_{\ell mn} = \{x \in [a, b] : a_\ell + \frac{1}{m} \leq f_n(x) \leq b_\ell - \frac{1}{m}\}$$
$$E_{\ell mN} = \bigcap_{n \geq N} F_{\ell mn}$$

closed?
(d) Show that

$$H_\ell = \bigcup_{m,N \in \mathbb{N}} E_{\ell mN}.$$

(e) Use (a) and Baire's Theorem (page 243) to deduce that some $E_{\ell mN}$ contains a subinterval of (α, β).
(f) Why does (e) contradict (b) and complete the proof that D_κ is nowhere dense?

23. Combine Exercises 18, 22, and Baire's Theorem to show that a Baire class 1 function has a dense set of continuity points.
24. Suppose that $g : [a, b] \to \mathbb{R}$ is differentiable.
(a) Prove that g' is of Baire class 1. [Hint: Extend g to a differentiable function defined on a larger interval and consider

$$f_n(x) = \frac{g(x + 1/n) - g(x)}{1/n}$$

for $x \in [a, b]$. Is $f_n(x)$ continuous? Does $f_n(x)$ converge pointwise to $g'(x)$ as $n \to \infty$?]
(b) Infer from Exercise 23 that *a derivative can not be everywhere discontinuous.* It must be continuous on a dense subset of its domain of definition.

25. Consider the characteristic functions $f(x)$ and $g(x)$ of the intervals $[1, 4]$ and $[2, 5]$. The derivatives f', g' exist almost everywhere. The integration by parts formula says that

$$\int_0^3 f(x)g'(x)\,dx = f(3)g(3) - f(0)g(0) - \int_0^3 f'(x)g(x)\,dx.$$

But both integrals are zero, while $f(3)g(3) - f(0)g(0) = 1$. Where is the error?

26. Let Ω be a set with a **transitive relation** $\preceq$. It satisfies the conditions that for all $\omega_1, \omega_2, \omega_3 \in \Omega$, $\omega_1 \preceq \omega_1$ and $\omega_1 \preceq \omega_2 \preceq \omega_3$ implies that $\omega_1 \preceq \omega_3$. A function $f : \Omega \to \mathbb{R}$ **converges to a limit** L with respect to Ω if, given any $\epsilon > 0$ there is an $\omega_0 \in \Omega$ such that $\omega_0 \preceq \omega$ implies $|f(\omega) - L| < \epsilon$. We write $\lim_\Omega f(\omega) = L$ to indicate this convergence. Observe that
- When $f(n) = a_n$ and $\mathbb{N}$ is given its standard order relation $\leq$, $\lim_{n\to\infty} a_n$ means the same thing as $\lim_{\mathbb{N}} f(n)$.
- When $\mathbb{R}$ is given its standard order relation $\leq$, $\lim_{t\to\infty} f(t)$ means the same thing as $\lim_{\mathbb{R}} f(t)$.
- Fix an $x \in \mathbb{R}$ and give $\mathbb{R}$ the new relation $t_1 \preceq t_2$ when $|t_2 - x| \leq |t_1 - x|$. Then $\lim_{t\to x} f(t)$ means the same thing as $\lim_{(\mathbb{R},\preceq)} f(t)$.

(a) Prove that limits are unique: if $\lim_\Omega f = L_1$ and $\lim_\Omega f = L_2$ then $L_1 = L_2$.
(b) Prove that existence of $\lim_\Omega f$ and $\lim_\Omega g$ imply that

$$\lim_\Omega (f + cg) = \lim_\Omega f + c \lim_\Omega g$$
$$\lim_\Omega (f \cdot g) = \lim_\Omega f \cdot \lim_\Omega g$$
$$\lim_\Omega (f/g) = \lim_\Omega f / \lim_\Omega g$$

where c is a constant and, in the quotient rule, $\lim_\Omega g \neq 0$, $g \neq 0$.
(c) Let Ω consist of all partition pairs (P, T); define $(P, T) \preceq (P', T')$ when P' is finer than P, mesh $P' \leq$ mesh P. Observe that $\preceq$ is transitive and that $\lim_\Omega R(f, P, T) = I$ means the same as $\lim_{\text{mesh } P\to 0} R(f, P, T)$ in the definition of the Riemann integral.
(d) Review the proof of Theorem 16 and use (b) to justify the fact that linearity of Riemann sums with respect to the integrand,

$$R(f + cg, P, T) = R(f, P, T) + cR(g, P, T),$$

actually does imply linearity of the integral with respect to the integrand.
(e) Formulate this limit definition for functions from Ω to a general metric space in place of $\mathbb{R}$.

27. Redefine Riemann and Darboux integrability using only dyadic partitions.
    (a) Prove that the integrals are unaffected.
    (b) Infer that Riemann's integrability criterion can be restated in terms of dyadic partitions.
28. In many calculus books, the definition of the integral is given as

$$\lim_{n\to\infty} \sum_{k=1}^{n} f(x_k^*) \frac{b-a}{n}$$

where x_k^* is the midpoint of the interval

$$[a + \frac{(k-1)(b-a)}{n},\ a + \frac{k(b-a)}{n}].$$

See Stewart's *Calculus with Early Transcendentals*, for example.
    (a) If f is continuous, show that the calculus book limit exists and equals the Riemann integral of f. [Hint: This is a one-liner.]
    (b) Show by example that the calculus style limit can exist for functions which are not Riemann integrable.
    (c) Infer that the calculus style definition of the integral is inadequate for real analysis.
29. Suppose that $Z \subset \mathbb{R}$. Prove that the following are equivalent.
    (i) Z is a zero set.
    (ii) For each $\epsilon > 0$ there is a countable covering of Z by closed intervals $[a_i, b_i]$ with total length $\sum b_i - a_i < \epsilon$.
    (iii) For each $\epsilon > 0$ there is a countable covering of Z by sets S_i with total diameter $\sum \text{diam } S_i < \epsilon$.
30. Prove that the interval $[0, 1]$ is not a zero set. [Hint: Be careful; this is not entirely trivial.]
31. The standard **middle-quarters Cantor set** is formed by removing the middle quarter from $[0, 1]$, then removing the middle quarter from each of the remaining two intervals, then removing the middle quarter from each of the remaining four intervals, and so on.
    (a) Prove that it is a zero set.
    (b) Formulate the natural definition of the middle β-Cantor set.
    (c) Is it also a zero set? Prove or disprove.

*32. Define a Cantor set by removing from $[0, 1]$ the middle interval of length $1/4$. From the remaining two intervals F^1 remove the middle intervals of length $1/16$. From the remaining four intervals F^2 remove the middle intervals of length $1/64$, and so on. At the n^{th} step in the construction F^n consists of 2^n subintervals of F^{n-1}.

    (a) Prove that $F = \bigcap F^n$ is a Cantor set but not a zero set. It is often referred to as a **fat Cantor set**.
    (b) Infer that being a zero set is not a topological property: if two sets are homeomorphic and one is a zero set, then the other need not be a zero set.

[Hint: To get a sense of this fat Cantor set, calculate the total length of the intervals which comprise its complement. See Figure 49 and Exercise 36.]

33. Consider the characteristic function of the dyadic rational numbers, $f(x) = 1$ if $x = k/2^n$ for some $k \in \mathbb{Z}$ and $n \in \mathbb{N}$, and $f(x) = 0$ otherwise.
    (a) What is its set of discontinuities?
    (b) At which points is its oscillation $\geq \kappa$?
    (c) Is it integrable? Explain, both by the Riemann-Lebesgue Theorem and directly from the definition.
    (d) Consider the **dyadic ruler function** $g(x) = 1/2^n$ if $x = k/2^n$ and $g(x) = 0$ otherwise. Graph it and answer the questions posed in (a), (b), (c).
34. (a) Prove that the characteristic function f of the middle-thirds Cantor set C is Riemann integrable but the characteristic function g of the fat Cantor set F (Exercise 32) is not.
    (b) Why is there a homeomorphism $h : [0, 1] \to [0, 1]$ sending C onto F?
    (c) Infer that the composite of Riemann integrable functions need not be Riemann integrable. How is this example related to Corollaries 26, 30 of the Riemann-Lebesgue Theorem? See also Exercise 36.

*35. Assume that $\psi : [a, b] \to \mathbb{R}$ is continuously differentiable. A **critical point** of ψ is an x such that $\psi'(x) = 0$. A **critical value** is a number y such that for at least one critical point x, $y = f(x)$.
    (a) Prove that the set of critical values is a zero set. (This is the Morse-Sard Theorem in dimension one.)
    (b) Generalize this to continuously differentiable functions $\mathbb{R} \to \mathbb{R}$.

*36. Let $F \subset [0, 1]$ be the fat Cantor set from Exercise 32, and define

$$\psi(x) = \int_0^x \text{dist}(t, F)\, dt$$

where $\text{dist}(t, F)$ refers to the minimum distance from t to F.

(a) Why is ψ a continuously differentiable homeomorphism from $[0, 1]$ onto $[0, L]$ where $L = \psi(1)$?
(b) What is the set of critical points of ψ? (See Exercise 35.)
(c) Why is $\psi(F)$ a Cantor set of zero measure?
(d) Let f be the characteristic function of $\psi(F)$. Why is f Riemann integrable but $f \circ \psi$ not?
(e) What is the relation of (d) to Exercise 34?

37. Generalizing Exercise 30 in Chapter 1, we say that $f : (a, b) \to \mathbb{R}$ has a **jump** discontinuity (or a discontinuity of the **first kind**) at $c \in (a, b)$ if

$$f(c^-) = \lim_{x \to c^-} f(x) \quad \text{and} \quad f(x^+) = \lim_{x \to c^+} f(x)$$

exist, but are either unequal or are unequal to $f(c)$. (The three quantities exist and are equal if and only if f is continuous at c.) An **oscillating** discontinuity (or a discontinuity of the **second kind**) is any non-jump discontinuity.
(a) Show that $f : \mathbb{R} \to \mathbb{R}$ has at most countably many jump discontinuities.
(b) Show that

$$f(x) = \begin{cases} \sin \dfrac{1}{x} & \text{if } x \geq 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

has an oscillating discontinuity at $x = 0$.
(c) Show that the characteristic function of the rationals, $\chi_{\mathbb{Q}}$, has an oscillating discontinuity at every point.

*38. Recall that $\mathcal{P}(S) = 2^S$ is the power set of S, the collection of all subsets of S and $\mathcal{R}$ is the set of Riemann integrable functions $f : [a, b] \to \mathbb{R}$.
(a) Prove that the cardinality of $\mathcal{R}$ is the same as the cardinality of $\mathcal{P}(\mathbb{R})$, which is greater than the cardinality of $\mathbb{R}$.
(b) Call two functions in $\mathcal{R}$ **integrally equivalent** if they differ only on a zero set. Prove that the collection of integral equivalence classes of $\mathcal{R}$ has the same cardinality as $\mathbb{R}$, namely $2^{\mathbb{N}}$.
(c) Is it better to count Riemann integrable functions or integral equivalence classes of Riemann integrable functions?
(d) Show that $f, g \in \mathcal{R}$ are integrally equivalent if and only if the integral of $|f - g|$ is zero.

39. Suppose that $\psi : [c, d] \to [a, b]$ is continuous and for every zero set $Z \subset [a, b]$, $\psi^{\text{pre}}(Z)$ is a zero set in $[c, d]$.

(a) If f is Riemann integrable, prove that $f \circ \psi$ is Riemann integrable.
(b) Derive Corollary 30 from (a).

40. Let $\psi(x) = x \sin 1/x$ for $0 < x \leq 1$ and $\psi(0) = 0$.
(a) If $f : [-1, 1] \to \mathbb{R}$ is Riemann integrable, prove that $f \circ \psi$ is Riemann integrable.
(b) What happens for $\psi(x) = \sqrt{x} \sin 1/x$?

*41. Assume that $\psi : [c, d] \to [a, b]$ is continuously differentiable.
(a) If the critical points of ψ form a zero set in $[c, d]$ and f is Riemann integrable on $[a, b]$ prove that $f \circ \psi$ is Riemann integrable on $[c, d]$.
(b) Conversely, prove that if $f \circ \psi$ is Riemann integrable for each Riemann integrable f on $[a, b]$, then the critical points of ψ form a zero set. [Hint: Think in terms of Exercise 35.]
(c) Prove (a) and (b) under the weaker assumption that ψ is continuously differentiable except at finitely many points of $[c, d]$.
(d) Derive part (a) of Exercise 36 from (c).
(e) Weaken the assumption further to ψ being continuously differentiable on an open subset of $[c, d]$ whose complement is a zero set.

The following assertion, to be proved in Chapter 6, is related to the preceding exercises. If $f : [a, b] \to \mathbb{R}$ satisfies a Lipschitz condition or is monotone then the set of points at which $f'(x)$ fails to exist is a zero set. Thus: "a Lipschitz function is differentiable almost everywhere," which is Rademacher's Theorem in dimension one, and a "monotone function is almost everywhere differentiable," which is the last theorem in Lebesgue's book, *Leçons sur l'intégration et la recherche des fonctions primitives*. See Theorem 39 and Corollary 41 in Chapter 6.

42. Set

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \sin \dfrac{\pi}{x} & \text{if } x > 0 \end{cases} \quad \text{and} \quad g(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0. \end{cases}$$

Prove that f has an antiderivative but g does not.

43. Show that any two antiderivatives of a function differ by a constant. [Hint: This is a one-liner.]

44. (a) Define the oscillation for a function from one metric space to another, $f : M \to N$.

(b) Is it true that f is continuous at a point if and only if its oscillation is zero there? Prove or disprove.

(c) Fix a number $\kappa > 0$. Is the set of points at which the oscillation of f is $\geq \kappa$ closed in M? Prove or disprove.

45. (a) Prove that the integral of the Zeno's staircase function described on page 161 is $2/3$.

(b) What about the Devil's staircase?

46. In the proof of Corollary 26 of the Riemann-Lebesgue Theorem, it is asserted that when ϕ is continuous the discontinuity set of $\phi \circ f$ is contained in the discontinuity set of f.

(a) Prove this.

(b) Give an example where the inclusion is not an equality.

(c) Find a sufficient condition on ϕ so that $\phi \circ f$ and f have equal discontinuity sets for all $f \in \mathcal{R}$

(d) Is your condition necessary too?

47. Assume that $f \in \mathcal{R}$ and for some $m > 0$, $|f(x)| \geq m$ for all $x \in [a, b]$. Prove that the reciprocal of f, $1/f(x)$, also belongs to $\mathcal{R}$. If $f \in \mathcal{R}$, $|f(x)| > 0$, but no $m > 0$ is an underbound for $|f|$, prove that the reciprocal of f is not Riemann integrable.

48. Corollary 26 to the Riemann-Lebesgue Theorem asserts that if $f \in \mathcal{R}$ and ϕ is continuous, then $\phi \circ f \in \mathcal{R}$. Show that piecewise continuity can not replace continuity. [Hint: Take f to be a ruler function and ϕ to be a characteristic function.]

**49. Assume that $f : [a, b] \to [c, d]$ is a Riemann integrable bijection. Is the inverse bijection also Riemann integrable? Prove or disprove. [Hint: The graph of f^{-1} is the same as the graph of f, viewed with the axes reversed. A function is Riemann integrable if and only if its graph is squeezed tightly between graphs of step functions. Draw a picture of this and rotate your paper by 90 degrees.]

50. If f, g are Riemann integrable on $[a, b]$, and $f(x) < g(x)$ for all $x \in [a, b]$, prove that $\int_a^b f(x)\,dx < \int_a^b g(x)\,dx$. (Note the *strict* inequality.)

51. Let $f : [a, b] \to \mathbb{R}$ be given. Prove or give counter-examples to the following assertions.

(a) $f \in \mathcal{R} \Rightarrow |f| \in \mathcal{R}$.

(b) $|f| \in \mathcal{R} \Rightarrow f \in \mathcal{R}$.

(c) $f \in \mathcal{R}$ and $|f(x)| \geq c > 0$ for all $x \Rightarrow 1/f \in \mathcal{R}$.

(d) $f \in \mathcal{R} \Rightarrow f^2 \in \mathcal{R}$.

(e) $f^2 \in \mathcal{R} \Rightarrow f \in \mathcal{R}$.

(f) $f^3 \in \mathcal{R} \Rightarrow f \in \mathcal{R}$.

(g) $f^2 \in \mathcal{R}$ and $f(x) \geq 0$ for all $x \Rightarrow f \in \mathcal{R}$.

[Here f^2 and f^3 refer to the functions $f(x) \cdot f(x)$ and $f(x) \cdot f(x) \cdot f(x)$, not the iterates.]

52. Given $f, g \in \mathcal{R}$, prove that $\max(f, g), \min(f, g) \in \mathcal{R}$, where $\max(f, g)(x) = \max(f(x), g(x))$ and $\min(f, g)(x) = \min(f(x), g(x))$.

53. Assume that $f, g : [0, 1] \to \mathbb{R}$ are Riemann integrable and $f(x) = g(x)$ except on the middle-thirds Cantor set C.

(a) Prove that f and g have the same integral.

(b) Is the same true if $f(x) = g(x)$ except for $x \in \mathbb{Q}$?

(c) How is this related to the fact that the characteristic function of $\mathbb{Q}$ is not Riemann integrable?

54. Prove that if $a_n \geq 0$ and $\sum a_n$ converges, then $\sum(\sqrt{a_n})/n$ converges.

55. (a) If $\sum a_n$ converges and (b_n) is monotonic and bounded, prove that $\sum a_n b_n$ converges.

(b) If the monotonicity condition is dropped, or replaced by the assumption that $\lim_{n\to\infty} b_n = 0$, find a counter-example to convergence of $\sum a_n b_n$.

56. Find an example of a series of positive terms that converges despite the fact that $\limsup_{n\to\infty} |a_{n+1}/a_n| = \infty$. Infer that ρ can not replace λ in the divergence half of the ratio test.

57. Prove that if the terms of a sequence decrease monotonically, $a_1 \geq a_2 \geq \ldots$, and converge to 0 then the series $\sum a_k$ converges if and only if the associated dyadic series

$$a_1 + 2a_2 + 4a_4 + 8a_8 + \cdots = \sum 2^k a_{2^k}$$

converges. (I call this the **block test** because it groups the terms of the series in blocks of length 2^{k-1}.)

58. Prove that $\sum 1/k \log(k)^p$ converges when $p > 1$ and diverges when $p \leq 1$. Here $k = 2, 3, \ldots$. [Hint: Integral test or block test.]

59. Concoct a series $\sum a_k$ such that $(-1)^k a_k > 0$, $a_k \to 0$, but the series diverges.

60. (a) Show that if a series has ratio lim sup ρ then it has exponential growth rate ρ. Infer that the ratio test is subordinate to the root test.

(b) Concoct a series such that the root test is conclusive but the ratio test is not. Infer that the root test has strictly wider scope than the ratio test.

61. Show that there is no simple comparison test for conditionally convergent series:
    (a) Find two series $\sum a_k$ and $\sum b_k$ such that $\sum b_k$ converges conditionally, $a_k/b_k \to 1$ as $k \to \infty$, and $\sum a_k$ diverges.
    (b) Why is this impossible if the series $\sum b_k$ is absolutely convergent?
62. An **infinite product** is an expression $\prod c_k$ where $c_k > 0$. The n^{th} **partial product** is $C_n = c_1 \cdots c_n$. If C_n converges to a limit $C \neq 0$ the product converges to C. Write $c_k = 1 + a_k$. If each $a_k \geq 0$ or each $a_k \leq 0$ prove that $\sum a_k$ converges if and only if $\prod c_k$ converges. [Hint: Take log's.]
63. Show that conditional convergence of the series $\sum a_k$ and the product $\prod(1 + a_k)$ can be unrelated to each other:
    (a) Set $a_k = (-1)^k/\sqrt{k}$. The series $\sum a_k$ converges but the corresponding product $\prod(1 + a_k)$ diverges.
    (b) Let $e_k = 0$ when k is odd and $e_k = 1$ when k is even. Set $b_k = e_k/k + (-1)^k/\sqrt{k}$. The series $\sum b_k$ diverges while the corresponding product $\prod(1 + b_k)$ converges.
64. Consider a series $\sum a_k$ and **rearrange** its terms by some bijection $\beta : \mathbb{N} \to \mathbb{N}$, forming a new series $\sum a_{\beta(k)}$. The rearranged series converges if and only if the partial sums $a_{\beta(1)} + \cdots + a_{\beta(n)}$ converge to a limit as $n \to \infty$.
    (a) Prove that every rearrangement of a convergent series of non-negative terms converges — and converges to the same sum as the original series.
    (b) Do the same for absolutely convergent series.

*65. Let $\sum a_k$ be given.
    (a) If $\sum a_k$ converges conditionally, prove that rearrangement totally alters its convergence in the sense that some rearrangements $\sum b_k$ of $\sum a_k$ diverge to $+\infty$, others diverge to $-\infty$, and still others converge to any given real number.
    (b) Infer that a series is absolutely convergent if and only if every rearrangement converges. (The fact that rearrangement radically alters conditional convergence shows that although finite addition is commutative, infinite addition (series) is not.)

**66. Suppose that $\sum a_k$ converges conditionally. If $\sum b_k$ is a rearrangement of $\sum a_k$, let Y be the set of subsequential limits of (B_n) where B_n is the n^{th} partial sum of $\sum b_k$. That is, $y \in Y$ if and only if some $B_{n_\ell} \to y$ as $\ell \to \infty$.

    (a) Prove that Y is closed and connected.
    (b) If Y is compact and non-empty, prove that $\sum b_k$ converges to Y in the sense that $d_H(Y_n, Y) \to 0$ as $n \to \infty$, where d_H is the Hausdorff metric on the space of compact subsets of $\mathbb{R}$ and Y_n is the closure of $\{B_m : m \geq n\}$. See Exercise 2.124.
    (c) Prove that each closed and connected subset of $\mathbb{R}$ is the set of subsequential limits of some rearrangement of $\sum a_k$.

The article, "The Remarkable Theorem of Lévy and Steinitz" by Peter Rosenthal in the American Math Monthly of April 1987 deals with some of these issues.

***67. Let V be a Banach space — a vector space with a norm such that V is complete with respect to the metric induced by the norm. (For example, $\mathbb{R}^m$ is a Banach space.) If $\sum v_k$ is a convergent series of vectors in V such that $\sum \|v_k\|$ diverges, what is the generalization of Exercise 66? In particular, is Y convex?

*68. Absolutely convergent series can be multiplied in a natural way, the result being their **Cauchy product**,

$$\left(\sum_{i=0}^{\infty} a_i\right)\left(\sum_{j=0}^{\infty} b_j\right) = \sum_{k=0}^{\infty} c_k$$

where $c_k = a_0 b_k + a_1 b_{k-1} + \cdots + a_k b_0$.
    (a) Prove that $\sum c_k$ converges absolutely.
    (b) Formulate some algebraic laws for such products (commutativity, distributivity, and so on.) Prove two of them.

[Hint for (a): Write the products $a_i b_j$ in an $\infty \times \infty$ matrix array M, and let A_n, B_n, C_n be the n^{th} partial sums of $\sum a_i$, $\sum b_j$, $\sum c_k$. You are asked to prove that $(\lim A_n)(\lim B_n) = \lim C_n$. The product of the limits is the limit of the products. The product $A_n B_n$ is the sum of all the $a_i b_j$ in the $n \times n$ corner submatrix of M and c_n is the sum of its anti-diagonal. Now estimate $A_n B_n - C_n$. Alternately, assume that $a_n, b_n \geq 0$ and draw a rectangle R with edges A, B. Observe that R is the union of rectangles R_{ij} with edges a_i, b_j.]

**69. With reference to Exercise 68,
    (a) Reduce the hypothesis that both series $\sum a_i$ and $\sum b_j$ are absolutely convergent to merely one being absolutely convergent and the other convergent. (Exercises 68 and 69(a) are known as **Mertens' Theorem**.)
    (b) Find an example to show that the Cauchy product of two conditionally convergent series may diverge.

70. The **Riemann ζ-function is defined to be $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ where $s > 1$. It is the sum of the p-series when $p = s$. Establish **Euler's product formula**,

$$\zeta(s) = \prod_{k=1}^{\infty} \frac{1}{1 - p_k^{-s}}$$

where p_k is the k^{th} prime number. Thus, $p_1 = 2$, $p_2 = 3$, and so on. Prove that the infinite product converges. [Hint: Each factor in the infinite product is the sum of a geometric series $1+p_k^{-s}+(p_k^{-s})^2+\ldots$. Replace each factor by its geometric series and write out the n^{th} partial product. Apply Mertens' Theorem, collect terms, and recall that any integer has a unique prime factorization.]

71. Invent a continuous function $f : \mathbb{R} \to \mathbb{R}$ whose improper integral is zero, but which is unbounded as $x \to -\infty$ and $x \to \infty$. [Hint: f is far from monotone.]

72. Assume that $f : \mathbb{R} \to \mathbb{R}$ and that restricted to any closed interval, f is bounded.
    (a) Formulate the concepts of conditional and absolute convergence of the improper Riemann integral of f.
    (b) Find an example that distinguishes them.

73. Let $f : [0, \infty) \to [0, \infty)$ and $\sum a_k$ be given. Assume that for all sufficiently large k and all $x \in [k, k+1)$, $f(x) \leq a_k$. Prove that divergence of the improper integral $\int_0^{\infty} f(x)\,dx$ implies divergence of $\sum a_k$.

***74. Let $f : \mathbb{R} \to \mathbb{R}$ be given. Assume that the square and cube of f are smooth. Is f smooth? That is, if $f \cdot f \in C^{\infty}$ and $f \cdot f \cdot f \in C^{\infty}$, does it follow that $f \in C^{\infty}$?

4

Function Spaces

1 Uniform Convergence and $C^0[a, b]$

Points converge to a limit if they get physically closer and closer to it. What about a sequence of functions? When do functions converge to a limit function? What should it mean that they get closer and closer to a limit function? The simplest idea is that a sequence of functions f_n converges to a limit function f if for each x, the values $f_n(x)$ converge to $f(x)$ as $n \to \infty$. This is called pointwise convergence: a sequence of functions $f_n : [a, b] \to \mathbb{R}$ **converges pointwise** to a limit function $f : [a, b] \to \mathbb{R}$ if for each $x \in [a, b]$,

$$\lim_{n \to \infty} f_n(x) = f(x).$$

The function f is the **pointwise limit** of the sequence (f_n) and we write

$$f_n \to f \quad \text{or} \quad \lim_{n \to \infty} f_n = f.$$

Note that the limit refers to $n \to \infty$, not to $x \to \infty$. The same definition applies to functions from one metric space to another.

The requirement of uniform convergence is stronger: the sequence of functions $f_n : [a, b] \to \mathbb{R}$ **converges uniformly** to the limit function $f : [a, b] \to \mathbb{R}$ if for each $\epsilon > 0$ there is an N such that for all $n \geq N$ and

all $x \in [a, b]$,

$$|f_n(x) - f(x)| < \epsilon. \tag{1}$$

The function f is the **uniform limit** of the sequence (f_n) and we write

$$f_n \rightrightarrows f \quad \text{or} \quad \operatorname*{unif\,lim}_{n\to\infty} f_n = f.$$

Your intuition about uniform convergence is crucial. Draw a tube V of vertical radius ϵ around the graph of f. For n large, the graph of f_n lies wholly in V. See Figure 83. Absorb this picture!

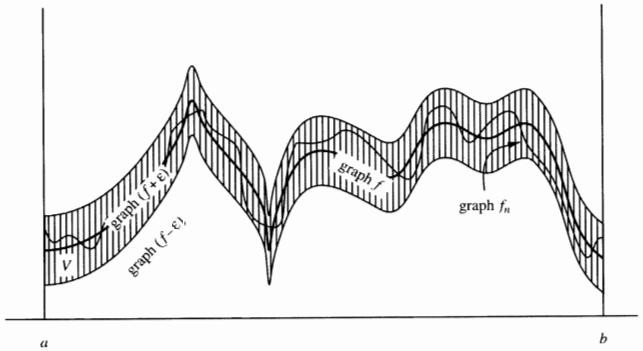


Figure 83 The graph of f_n is contained in the ϵ-tube around the graph of f.

It is clear that uniform convergence implies pointwise convergence. The difference between the two definitions is apparent in the following standard example.

Example Define $f_n : (0, 1) \to \mathbb{R}$ by $f_n(x) = x^n$. For each $x \in (0, 1)$ it is clear that $f_n(x) \to 0$. The functions converge pointwise to the zero function as $n \to \infty$. They do not converge uniformly: for take $\epsilon = 1/10$. The point $x_n = \sqrt[n]{1/2}$ is sent by f_n to $1/2$ and thus not all points x satisfy (1) when n is large. The graph of f_n fails to lie in the ϵ-tube V. See Figure 84.

The lesson to draw is that pointwise convergence of a sequence of functions is frequently too weak a concept. Gravitating toward uniform convergence we ask the natural question:

Which properties of functions are preserved under uniform convergence?

The answers are found in Theorem 1, Exercise 4, Theorem 6, and Theorem 9. Uniform limits preserve continuity, uniform continuity, integrability, and — with an additional hypothesis — differentiability.

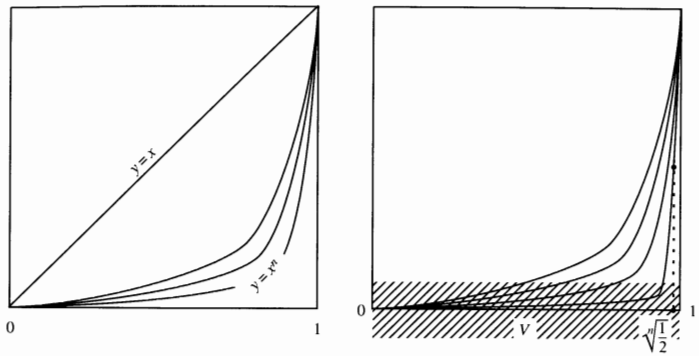


Figure 84 Non-uniform, pointwise convergence.

1 Theorem *If $f_n \rightrightarrows f$ and each f_n is continuous at x_0, then f is continuous at x_0. In other words, the uniform limit of continuous functions is continuous.*

Proof For simplicity, assume that the functions have domain $[a, b]$ and target $\mathbb{R}$. (See also Section 8 and Exercise 2.) Let $\epsilon > 0$ and $x_0 \in [a, b]$ be given. There is an N such that for all $n \geq N$ and all $x \in [a, b]$,

$$|f_n(x) - f(x)| < \frac{\epsilon}{3}.$$

The function f_N is continuous at x_0 and so there is a $\delta > 0$ such that $|x - x_0| < \delta$ implies

$$|f_N(x) - f_N(x_0)| < \frac{\epsilon}{3}.$$

If $|x - x_0| < \delta$ then

$$\begin{aligned} |f(x) - f(x_0)| &\leq |f(x) - f_N(x)| + |f_N(x) - f_N(x_0)| + |f_N(x_0) - f(x_0)| \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Thus f is continuous at $x_0 \in [a, b]$. □

Without uniform convergence, the theorem fails. For example, define $f_n : [0, 1] \to \mathbb{R}$ as before, $f_n(x) = x^n$. Then $f_n(x)$ converges pointwise to the function

$$f(x) = \begin{cases} 0 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x = 1. \end{cases}$$

The function f is not continuous and the convergence is not uniform. What about the converse? If the limit and the functions are continuous, does pointwise convergence imply uniform convergence? The answer is "no," as is shown by x^n on $(0, 1)$. But what if the functions have a compact domain of definition, $[a, b]$? The answer is still "no."

Example John Kelley refers to this as the **growing steeple**,

$$f_n(x) = \begin{cases} n^2 x & \text{if } 0 \le x \le \frac{1}{n} \\ 2n - n^2 x & \text{if } \frac{1}{n} \le x \le \frac{2}{n} \\ 0 & \text{if } \frac{2}{n} \le x \le 1. \end{cases}$$

See Figure 85.

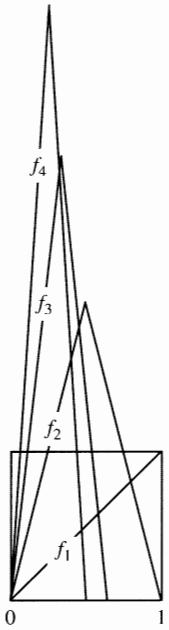


Figure 85 The sequence of functions converges pointwise to the zero function, but not uniformly.

Then $\lim_{n\to\infty} f_n(x) = 0$ for each x, and f_n converges pointwise to the function $f = 0$. Even if the functions have compact domain of definition, and are uniformly bounded and uniformly continuous, pointwise convergence does not imply uniform convergence. For an example, just multiply the growing steeple functions by $1/n$.

The *natural* way to view uniform convergence is in a function space. Let $C_b = C_b([a, b], \mathbb{R})$ denote the set of all bounded functions $[a, b] \to \mathbb{R}$.

The elements of C_b are functions f, g, etc. Each is bounded. Define the **sup norm** on C_b as

$$\|f\| = \sup\{|f(x)| : x \in [a, b]\}.$$

The sup norm satisfies the norm axioms discussed in Chapter 1, page 27.

$\|f\| \ge 0$ and $\|f\| = 0$ if and only if $f = 0$.
$\|cf\| = |c|\,\|f\|$
$\|f + g\| \le \|f\| + \|g\|$.

As we observed in Chapter 2, any norm defines a metric. In the case at hand,

$$d(f, g) = \sup\{|f(x) - g(x)| : x \in [a, b]\}$$

is the corresponding metric on C_b. See Figure 86. To distinguish the norm $\|f\| = \sup|f(x)|$ from other norms on C_b we sometimes write $\|f\|_{\sup}$ for the sup norm.

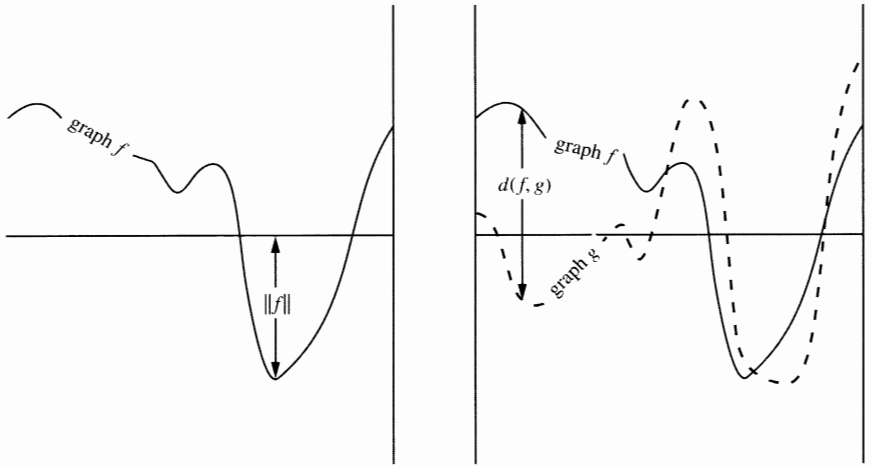


Figure 86 The sup-norm of f and the sup-distance between the functions f and g.

The thing to remember is that C_b is a metric space whose *elements* are functions. Ponder this.

2 Theorem *Convergence with respect to the sup-metric d is equivalent to uniform convergence.*

Proof If $d(f_n, f) \to 0$ then $\sup\{|f_n x - fx| : x \in [a, b]\} \to 0$, so $f_n \rightrightarrows f$, and conversely. □

3 Theorem *C_b is a complete metric space.*

Proof Let (f_n) be a Cauchy sequence in C_b. For each individual $x_0 \in [a, b]$ the values $f_n(x_0)$ form a Cauchy sequence in $\mathbb{R}$ since

$$|f_n(x_0) - f_m(x_0)| \leq \sup\{|f_n(x) - f_m(x)| : x \in [a, b]\} = d(f_n, f_m).$$

Thus, for each $x \in [a, b]$,

$$\lim_{n\to\infty} f_n(x)$$

exists. Define this limit to be $f(x)$. It is clear that f_n converges pointwise to f. In fact, the convergence is uniform. For let $\epsilon > 0$ be given. There exists N such that $m, n \geq N$ imply

$$d(f_n, f_m) < \frac{\epsilon}{2}.$$

Also, for each $x \in [a, b]$ there exists an $m = m(x) \geq N$ such that

$$|f_m(x) - f(x)| < \frac{\epsilon}{2}.$$

If $n \geq N$ and $x \in [a, b]$ then

$$\begin{aligned} |f_n(x) - f(x)| &\leq \left|f_n(x) - f_{m(x)}(x)\right| + \left|f_{m(x)}(x) - f(x)\right| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

Hence $f_n \rightrightarrows f$. The function f is bounded. For f_N is bounded and for all x, $|f_N(x) - f(x)| < \epsilon$. Thus $f \in C_b$. By Theorem 2, uniform convergence implies d-convergence, $d(f_n, f) \to 0$, and the Cauchy sequence (f_n) converges to a limit in the metric space C_b. □

The preceding proof is subtle. The uniform inequality $d(f_n, f) < \epsilon$ is derived by non-uniform means: for each x we make a separate estimate using an $m(x)$ depending non-uniformly on x. It is a case of the ends justifying the means.

Let $C^0 = C^0([a, b], \mathbb{R})$ denote the set of continuous functions $[a, b] \to \mathbb{R}$. Each $f \in C^0$ belongs to C_b since a continuous function defined on a compact domain is bounded. That is, $C^0 \subset C_b$.

4 Corollary *C^0 is a closed subset of C_b. It is a complete metric space.*

Proof Theorem 1 implies that a limit in C_b of a sequence of functions in C^0 lies in C^0. That is, C^0 is closed in C_b. A closed subset of a complete space is complete. □

Just as it is reasonable to discuss the convergence of a sequence of functions we can also discuss the convergence of a series of functions, $\sum f_k$. Merely consider the n^{th} partial sum

$$F_n(x) = \sum_{k=0}^{n} f_k(x).$$

It is a function. If the sequence of functions (F_n) converges to a limit function F then the series converges, and we write

$$F(x) = \sum_{k=0}^{\infty} f_k(x).$$

If the sequence of partial sums converges uniformly, then so does the series. If the series of absolute values $\sum |f_k(x)|$ converges, then the series $\sum f_k$ converges absolutely.

5 Weierstrass M-test *If $\sum M_k$ is a convergent series of constants and if $f_k \in C_b$ satisfies $\|f_k\| \leq M_k$ for all k, then $\sum f_k$ converges uniformly and absolutely.*

Proof If $n > m$ then the partial sums of the series of absolute values telescope as

$$\begin{aligned} d(F_n, F_m) &\leq d(F_n, F_{n-1}) + \cdots + d(F_{m+1}, F_m) \\ &= \sum_{k=m+1}^{n} \|f_k\| \leq \sum_{k=m+1}^{n} M_k. \end{aligned}$$

Since $\sum M_k$ converges, the last sum is $< \epsilon$ when m, n are large. Thus (F_n) is Cauchy in C_b, and by Theorem 3 it converges uniformly. □

Next we ask how integrals and derivatives behave with respect to uniform convergence. Integrals behave better than derivatives.

6 Theorem *The uniform limit of Riemann integrable functions is Riemann integrable, and the limit of the integrals is the integral of the limit,*

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b \operatorname*{unif\,lim}_{n\to\infty} f_n(x)\,dx.$$

In other words, $\mathcal{R}$ is a closed subset of C_b and the integral functional $f \mapsto \int_a^b f(x)\,dx$ is a continuous map from $\mathcal{R}$ to $\mathbb{R}$. This extends the regularity hierarchy to

$$C_b \supset \mathcal{R} \supset C^0 \supset C^1 \supset \cdots \supset C^\infty \supset C^\omega.$$

Theorem 6 gives the simplest condition under which the operations of taking limits and integrals commute.

Proof Let $f_n \in \mathcal{R}$ be given and assume that $f_n \rightrightarrows f$ as $n \to \infty$. By the Riemann-Lebesgue Theorem, f_n is bounded and there is a zero set Z_n such that f_n is continuous at each $x \in [a,b]\backslash Z_n$. Theorem 1 implies that f is continuous at each $x \in [a,b]\backslash\bigcup Z_n$, while Theorem 3 implies that f is bounded. Since $\bigcup Z_n$ is a zero set, the Riemann-Lebesgue Theorem implies that $f \in \mathcal{R}$. Finally

$$\left|\int_a^b f(x)\,dx - \int_a^b f_n(x)\,dx\right| = \left|\int_a^b f(x) - f_n(x)\,dx\right|$$
$$\leq \int_a^b |f(x) - f_n(x)|\,dx \leq d(f, f_n)(b-a) \to 0$$

as $n \to \infty$. Hence the integral of the limit is the limit of the integrals. □

7 Corollary *If $f_n \in \mathcal{R}$ and $f_n \rightrightarrows f$ then the indefinite integrals converge uniformly,*

$$\int_a^x f_n(t)\,dt \rightrightarrows \int_a^x f(t)\,dt.$$

Proof As above,

$$\left|\int_a^x f(t)\,dt - \int_a^x f_n(t)\,dt\right| \leq d(f_n, f)(x-a) \leq d(f_n, f)(b-a) \to 0$$

when $n \to \infty$. □

8 Term by Term Integration Theorem *A uniformly convergent series of integrable functions $\sum f_k$ can be integrated term by term in the sense that*

$$\int_a^b \sum_{k=0}^{\infty} f_k(x)\,dx = \sum_{k=0}^{\infty} \int_a^b f_k(x)\,dx.$$

Proof The sequence of partial sums F_n converges uniformly to $\sum f_k$. Each F_n belongs to $\mathcal{R}$ since it is the finite sum of members of $\mathcal{R}$. According to Theorem 6,

$$\sum_{k=0}^{n} \int_a^b f_k(x)\,dx = \int_a^b F_n(x)\,dx \to \int_a^b \sum_{k=0}^{\infty} f_k(x)\,dx.$$

This shows that the series $\sum \int_a^b f_k(x)\,dx$ converges to $\int_a^b \sum f_k(x)\,dx$. □

9 Theorem *The uniform limit of a sequence of differentiable functions is differentiable provided that the sequence of derivatives also converges uniformly.*

Proof We suppose that $f_n : [a,b] \to \mathbb{R}$ is differentiable for each n and that $f_n \rightrightarrows f$ as $n \to \infty$. Also we assume that $f_n' \rightrightarrows g$ for some function g. Then we show that f is differentiable and in fact $f' = g$.

We first prove the theorem with a major loss of generality: we assume that each f_n' is continuous. Then $f_n', g \in \mathcal{R}$ and we can apply the fundamental theorem of calculus and Corollary 7 to write

$$f_n(x) \quad = \quad f_n(a) + \int_a^x f_n'(t)\,dt \quad \rightrightarrows \quad f(a) + \int_a^x g(t)\,dt.$$

Since $f_n \rightrightarrows f$ we see that $f(x) = f(a) + \int_a^x g(t)\,dt$ and, again by the fundamental theorem of calculus, $f' = g$.

In the general case the proof is harder. Fix some $x \in [a,b]$ and define

$$\phi_n(t) = \begin{cases} \dfrac{f_n(t) - f_n(x)}{t - x} & \text{if } t \neq x \\ f_n'(x) & \text{if } t = x \end{cases}$$

$$\phi(t) = \begin{cases} \dfrac{f(t) - f(x)}{t - x} & \text{if } t \neq x \\ g(x) & \text{if } t = x. \end{cases}$$

Each function ϕ_n is continuous since $\phi_n(t)$ converges to $f_n'(x)$ as $t \to x$. Also it is clear that ϕ_n converges pointwise to ϕ as $n \to \infty$. We claim the convergence is uniform. For any m, n the Mean Value Theorem applied to the function $f_m - f_n$ gives

$$\phi_m(t) - \phi_n(t) = \frac{(f_m(t) - f_n(t)) - (f_m(x) - f_n(x))}{t - x} = f_m'(\theta) - f_n'(\theta)$$

for some θ between t and x. Since $f_n' \rightrightarrows g$ the difference $f_m' - f_n'$ tends uniformly to 0 as $m, n \to \infty$. Thus (ϕ_n) is Cauchy in C^0. Since C^0 is complete, ϕ_n converges uniformly to a limit function ψ, and ψ is continuous. As already remarked, the pointwise limit of ϕ_n is ϕ, and so $\psi = \phi$. Continuity of $\psi = \phi$ implies that $g(x) = f'(x)$. □

10 Term by Term Differentiation Theorem *A uniformly convergent series of differentiable functions can be differentiated term by term, provided that the derivative series converges uniformly,*

$$\left(\sum_{k=0}^{\infty} f_k(x)\right)' = \sum_{k=0}^{\infty} f_k'(x).$$

Proof Apply Theorem 9 to the sequence of partial sums. □

Note that Theorem 9 fails if we forget to assume the derivatives converge. For example, consider the sequence of functions $f_n : [-1, 1] \to \mathbb{R}$ defined by

$$f_n(x) = \sqrt{x^2 + \frac{1}{n}}.$$

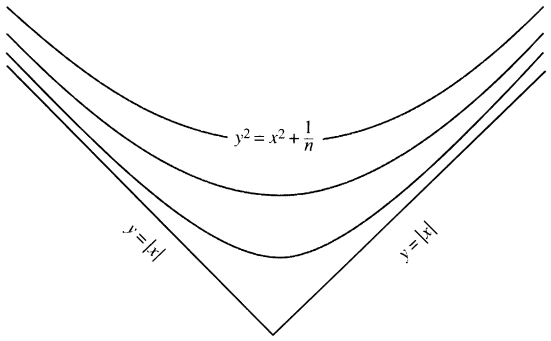


Figure 87 The uniform limit of differentiable functions need not be differentiable.

See Figure 87. The functions converge uniformly to $f(x) = |x|$, a non-differentiable function. The derivatives converge pointwise but not uniformly. Worse examples are easy to imagine. In fact, a sequence of everywhere differentiable functions can converge uniformly to a nowhere differentiable function. See Sections 4 and 7. It is one of the miracles of the complex numbers that a uniform limit of complex differentiable functions is complex differentiable, and automatically the sequence of derivatives converges uniformly to a limit. Real and complex analysis diverge radically on this point.

2 Power Series

As another application of the Weierstrass M-test we say a little more about the power series $\sum c_k x^k$. A power series is a special type of series of functions, the functions being constant multiples of powers of x. As explained in Section 3 of Chapter 3, its radius of convergence is

$$R = \frac{1}{\limsup_{k\to\infty} \sqrt[k]{c_k}}.$$

Its interval of convergence is $(-R, R)$. If $x \in (-R, R)$, the series converges and defines a function $f(x) = \sum c_k x^k$, while if $x \notin [-R, R]$, the series diverges. More is true on compact subintervals of $(-R, R)$.

11 Theorem *If $r < R$, then the power series converges uniformly and absolutely on the interval $[-r, r]$.*

Proof Choose β, $r < \beta < R$. For all large k, $\sqrt[k]{|c_k|} < 1/\beta$ since $\beta < R$. Thus, if $|x| \le r$ then

$$|c_k x^k| \le \left(\frac{r}{\beta}\right)^k.$$

These are terms in a convergent geometric series and according to the M-test $\sum c_k x^k$ converges uniformly when $x \in [-r, r]$. □

12 Theorem *A power series can be integrated and differentiated term by term on its interval of convergence.*

For $f(x) = \sum c_k x^k$ and $|x| < R$ this means

$$\int_0^x f(t)\,dt = \sum_{k=0}^{\infty} \frac{c_k}{k+1} x^{k+1} \quad \text{and} \quad f'(x) = \sum_{k=1}^{\infty} k c_k x^{k-1}.$$

Proof The radius of convergence of the integral series is determined by the exponential growth rate of its coefficients,

$$\limsup_{k\to\infty} \sqrt[k]{\left|\frac{c_{k-1}}{k}\right|} = \limsup_{k\to\infty} \left(|c_{k-1}|^{1/(k-1)}\right)^{(k-1)/k} \left(\frac{1}{k}\right)^{1/k}.$$

Since $(k-1)/k \to 1$ and $k^{-1/k} \to 1$ as $k \to \infty$, we see that the integral series has the same radius of convergence R as the original series. According to Theorem 8, term by term integration is valid when the series converges

uniformly, and by Theorem 11, the integral series does converge uniformly on any interval $[-r, r] \subset (-R, R)$.

A similar calculation for the derivative series shows that its radius of convergence too is R. Term by term differentiation is valid provided the series and the derivative series converge uniformly. Since the radius of convergence of the derivative series is R, the derivative series does converge uniformly on any $[-r, r] \subset (-R, R)$. □

13 Theorem *Analytic functions are smooth, $C^\omega \subset C^\infty$.*

Proof An analytic function f is defined by a convergent power series. According to Theorem 12, the derivative of f is given by a convergent power series with the same radius of convergence, so repeated differentiation is valid, and we see that f is indeed smooth. □

The general smooth function is not analytic, as is shown by the example

$$e(x) = \begin{cases} e^{-1/x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

on page 149. Near $x = 0$, $e(x)$ can not be expressed as a convergent power series.

Power series provide the clean and unambiguous way to define functions, especially trigonometric functions. The usual definitions of sine, cosine, etc. involve angles and circular arc length, and these concepts seem less fundamental than the functions being defined. To avoid circular reasoning, as it were, we declare that by definition

$$\exp x = \sum_{k\neq 0}^{\infty} \frac{x^k}{k!} \qquad \sin x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!} \qquad \cos x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!}.$$

We then must prove that these functions have the properties we know and love from calculus. All three series are easily seen to have radius of convergence $R = \infty$. Theorem 12 justifies term by term differentiation, yielding the usual formulas,

$$\exp'(x) = \exp x \qquad \sin'(x) = \cos x \qquad \cos'(x) = -\sin x.$$

The logarithm has already been defined as the indefinite integral $\int_1^x 1/t\, dt$. We claim that if $|x| < 1$ then $\log(1 + x)$ is given as the power series

$$\log(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1} x^k}{k}.$$

To check this, we merely note that its derivative is the sum of a geometric series,

$$(\log(1+x))' = \frac{1}{x+1} = \frac{1}{1-(-x)} = \sum_{k=0}^{\infty} (-x)^k = \sum_{k=0}^{\infty} (-1)^k x^k.$$

The last is a power series with radius of convergence 1. Since term by term integration of a power series inside its radius of convergence is legal, we integrate both sides of the equation and get the series expression for $\log(1 + x)$ as claimed.

3 Compactness and Equicontinuity in C^0

The Heine-Borel theorem states that a closed and bounded set in $\mathbb{R}^m$ is compact. On the other hand, closed and bounded sets in C^0 are rarely compact. Consider for example the closed unit ball

$$\mathcal{B} = \{f \in C^0([0,1], \mathbb{R}) : \|f\| \leq 1\}.$$

To see that $\mathcal{B}$ is not compact we look again at the sequence $f_n(x) = x^n$. It lies in $\mathcal{B}$. Does it have a subsequence that converges (with respect to the metric d of C^0) to a limit in C^0? No. For if f_{n_k} converges to f in C^0 then $f(x) = \lim_{k\to\infty} f_{n_k}(x)$. Thus $f(x) = 0$ if $x < 1$ and $f(1) = 1$, but this function f does not belong to C^0. The cause of the problem is the fact that C^0 is infinite-dimensional. In fact it can be shown that if V is a vector space with a norm then its closed unit ball is compact if and only if the space is finite-dimensional. The proof is not especially hard.

Nevertheless, we want to have theorems that guarantee certain closed and bounded subsets of C^0 are compact. For we want to extract a convergent subsequence of functions from a given sequence of functions. The simple condition that lets us go ahead is equicontinuity. A sequence of functions (f_n) in C^0 is **equicontinuous** if

$$\begin{aligned} &\forall \epsilon > 0 \ \ \exists \delta > 0 \text{ such that} \\ &|s - t| < \delta \text{ and } n \in \mathbb{N} \quad \Rightarrow \quad |f_n(s) - f_n(t)| < \epsilon. \end{aligned}$$

The functions f_n are *equally continuous*. The δ depends on ϵ but it does not depend on n. Roughly speaking, the graphs of all the f_n are similar. For total clarity, the concept might better be labeled uniform equicontinuity, in contrast to **pointwise equicontinuity**, which requires

$$\begin{aligned} &\forall \epsilon > 0 \text{ and } \forall x \in [a,b] \ \ \exists \delta > 0 \text{ such that} \\ &|x - t| < \delta \text{ and } n \in \mathbb{N} \quad \Rightarrow \quad |f_n(x) - f_n(t)| < \epsilon. \end{aligned}$$

The definitions work equally well for sets of functions, not only sequences of functions. The set $\mathcal{E} \subset C^0$ is equicontinuous if

$$\forall \epsilon > 0 \ \ \exists \delta > 0 \text{ such that}$$
$$|s - t| < \delta \text{ and } f \in \mathcal{E} \quad \Rightarrow \quad |f(s) - f(t)| < \epsilon.$$

The crucial point is that δ does not depend on the particular $f \in \mathcal{E}$. It is valid for all $f \in \mathcal{E}$ simultaneously. To picture equicontinuity of a family $\mathcal{E}$, imagine the graphs. Their shapes are uniformly controlled. Note that any finite number of continuous functions $[a, b] \to \mathbb{R}$ forms an equicontinuous family so Figures 88 and 89 are only suggestive.

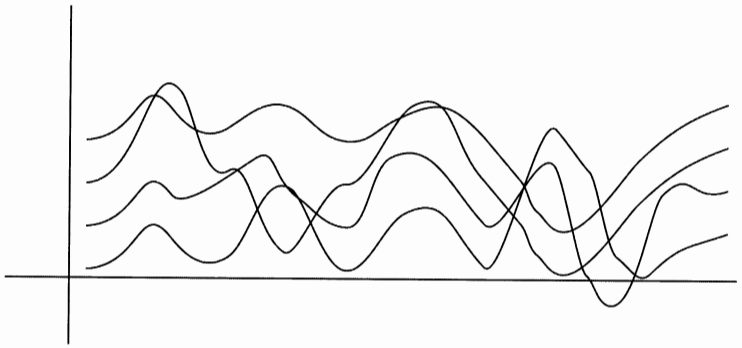


Figure 88 Equicontinuity.

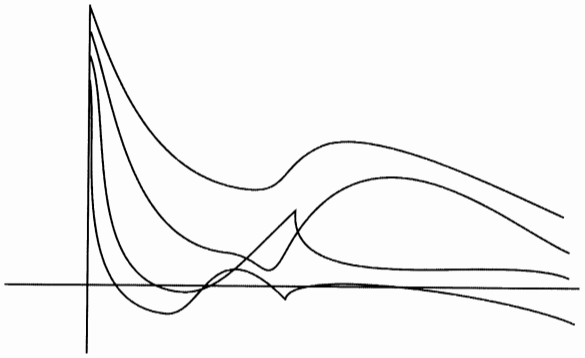


Figure 89 Non-equicontinuity.

The basic theorem about equicontinuity is the

14 Arzela-Ascoli Theorem *Any bounded equicontinuous sequence of functions in* $C^0([a, b], \mathbb{R})$ *has a uniformly convergent subsequence.*

Think of this as a compactness result. If (f_n) is the sequence of equicontinuous functions, the theorem amounts to asserting that the closure of the set $\{f_n : n \in \mathbb{N}\}$ is compact. Any compact metric space serves just as well as $[a, b]$, and the target space $\mathbb{R}$ can also be more general. See Section 8.

Proof $[a, b]$ has a countable dense subset $D = \{d_1, d_2, \dots\}$. For instance we could take $D = \mathbb{Q} \cap [a, b]$. Boundedness of (f_n) means that for some constant M, all $x \in [a, b]$, and all $n \in \mathbb{N}$, $|f_n(x)| \leq M$. Thus $(f_n(d_1))$ is a bounded sequence of real numbers. Bolzano-Weierstrass implies that some subsequence of it converges to a limit in $\mathbb{R}$, say

$$f_{1,k}(d_1) \to y_1 \text{ as } k \to \infty.$$

The subsequence $f_{1,k}$ evaluated at the point d_2 is also a bounded sequence in $\mathbb{R}$, and there exists a sub-subsequence $f_{2,k}$ such that $f_{2,k}(d_2)$ converges to a limit in $\mathbb{R}$, say $f_{2,k}(d_2) \to y_2$ as $k \to \infty$. The sub-subsequence evaluated at d_1 still converges to y_1. Continuing in this way gives a nested family of subsequences $f_{m,k}$ such that

$$(f_{m,k}) \text{ is a subsequence of } (f_{m-1,k})$$
$$j \leq m \quad \Rightarrow \quad f_{m,k}(d_j) \to y_j \text{ as } k \to \infty.$$

Choose $k(m) \geq m$ large enough that if $j \leq m$ and $k(m) \leq k$ then

$$\left|f_{m,k}(d_j) - y_j\right| < \frac{1}{m}.$$

The superdiagonal subsequence $g_m(x) = f_{m,k(m)}(x)$ converges to a limit at each point $x \in D$. We claim that $g_m(x)$ also converges at the other points $x \in [a, b]$ and that the convergence is uniform. It suffices to show that (g_m) is a Cauchy sequence in C^0.

Let $\epsilon > 0$ be given. Equicontinuity gives a $\delta > 0$ such that for all $s, t \in [a, b]$,

$$|s - t| < \delta \quad \Rightarrow \quad |g_m(s) - g_m(t)| < \frac{\epsilon}{3}.$$

Choose J large enough that every $x \in [a, b]$ lies in the δ-neighborhood of some d_j with $j \leq J$. Since D is dense and $[a, b]$ is compact, this is possible. See Exercise 19. Since $\{d_1, \dots, d_J\}$ is a finite set and $g_m(d_j)$ converges for each d_j, there is an N such that for all $\ell, m \geq N$ and all $j \leq J$,

$$\left|g_m(d_j) - g_\ell(d_j)\right| < \frac{\epsilon}{3}.$$

If $\ell, m \geq N$ and $x \in [a, b]$, choose d_j with $|d_j - x| < \delta$ and $j \leq J$. Then

$$\begin{aligned} |g_m(x) - g_\ell(x)| &\leq |g_m(x) - g_m(d_j)| + |g_m(d_j) - g_\ell(d_j)| + |g_\ell(d_j) - g_\ell(x)| \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Hence (g_m) is Cauchy in C^0, it converges in C^0, and the proof is complete. □

Part of the preceding development can be isolated as the

15 Arzela-Ascoli Propagation Theorem *Pointwise convergence of an equicontinuous sequence of functions on a dense subset of the domain* ***propagates*** *to uniform convergence on the whole domain.*

Proof This is the $\epsilon/3$ part of the proof. □

The example cited over and over again in the equicontinuity world is the following:

16 Corollary *Assume that $f_n : [a, b] \to \mathbb{R}$ is a sequence of differentiable functions whose derivatives are uniformly bounded. If for some x_0, $f_n(x_0)$ is bounded as $n \to \infty$, then the sequence (f_n) has a subsequence that converges uniformly on $[a, b]$.*

Proof Let M be a bound for the derivatives $|f_n'(x)|$, valid for all $n \in \mathbb{N}$ and $x \in [a, b]$. Equicontinuity of (f_n) follows from the Mean Value Theorem:

$$|s - t| < \delta \quad \Rightarrow \quad |f_n(s) - f_n(t)| = |f_n'(\theta)|\,|s - t| \leq M\delta$$

for some θ between s and t. Thus, given $\epsilon > 0$, the choice $\delta = \epsilon/(M + 1)$ shows that (f_n) is equicontinuous.

Let C be a bound for $|f_n(x_0)|$, valid for all $n \in \mathbb{N}$. Then

$$\begin{aligned} |f_n(x)| &\leq |f_n(x) - f_n(x_0)| + |f_n(x_0)| \leq M\,|x - x_0| + C \\ &\leq M\,|b - a| + C \end{aligned}$$

shows that the sequence (f_n) is bounded in C^0. The Arzela-Ascoli theorem then supplies the uniformly convergent subsequence. □

Two other consequences of the same type are fundamental theorems in the fields of ordinary differential equations and complex variables.

(a) A sequence of solutions to a continuous ordinary differential equation in $\mathbb{R}^m$ has a subsequence that converges to a limit, and that limit is also a solution of the ODE.

(b) A sequence of complex analytic functions that converges pointwise, converges uniformly (on compact subsets of the domain of definition) and the limit is complex analytic.

Finally we give a topological interpretation of the Arzela-Ascoli theorem.

17 Heine-Borel Theorem in a function space *A subset $\mathcal{E} \subset C^0$ is compact if and only if it is closed, bounded, and equicontinuous.*

Proof Assume that $\mathcal{E}$ is compact. By Theorem 2.56, it is closed and totally bounded: given $\epsilon > 0$ there is a finite covering of $\mathcal{E}$ by neighborhoods in C^0 that have radius $\epsilon/3$, say $\mathcal{N}_{\epsilon/3}(f_k)$, with $k = 1, \ldots, n$. Each f_k is uniformly continuous so there is a $\delta > 0$ such that

$$|s - t| < \delta \quad \Rightarrow \quad |f_k(s) - f_k(t)| < \frac{\epsilon}{3}.$$

If $f \in \mathcal{E}$ then for some k, $f \in \mathcal{N}_{\epsilon/3}(f_k)$, and $|s - t| < \delta$ implies

$$\begin{aligned} |f(s) - f(t)| &\leq |f(s) - f_k(s)| + |f_k(s) - f_k(t)| + |f_k(t) - f(t)| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \end{aligned}$$

Thus $\mathcal{E}$ is equicontinuous.

Conversely, assume that $\mathcal{E}$ is closed, bounded, and equicontinuous. If (f_n) is a sequence in $\mathcal{E}$ then by the Arzela-Ascoli theorem, some subsequence (f_{n_k}) converges uniformly to a limit. The limit lies in $\mathcal{E}$ since $\mathcal{E}$ is closed. Thus $\mathcal{E}$ is compact. □

4 Uniform Approximation in C^0

Given a continuous but nondifferentiable function f, we often want to make it smoother by a small perturbation. We want to approximate f in C^0 by a smooth function g. The ultimately smooth function is a polynomial, and the first thing we prove is a polynomial approximation result.

18 Weierstrass Approximation Theorem *The set of polynomials is dense in $C^0([a, b], \mathbb{R})$.*

Density means that for each $f \in C^0$ and each $\epsilon > 0$ there is a polynomial function $p(x)$ such that for all $x \in [a, b]$,

$$|f(x) - p(x)| < \epsilon.$$

There are several proofs of this theorem, and although they appear quite

different from each other, they share a common thread: the approximating function is built from f by sampling the values of f and recombining them in some nice way. It is no loss of generality to assume that the interval $[a, b]$ is $[0, 1]$. We do so.

Proof #1 For each $n \in \mathbb{N}$, consider the sum

$$p_n(x) = \sum_{k=0}^{n} \binom{n}{k} c_k x^k (1-x)^{n-k},$$

where $c_k = f(k/n)$ and $\binom{n}{k}$ is the binomial coefficient $n!/k!(n-k)!$. Clearly p_n is a polynomial. It is called a **Bernstein polynomial**. We claim that the n^{th} Bernstein polynomial converges uniformly to f as $n \to \infty$. The proof relies on two formulas about how the functions

$$r_k(x) = \binom{n}{k} x^k (1-x)^{n-k}$$

shown in Figure 90 behave. They are

$$\sum_{k=0}^{n} r_k(x) = 1 \tag{2}$$

$$\sum_{k=0}^{n} (k - nx)^2 r_k(x) = nx(1-x). \tag{3}$$

In terms of the functions r_k we write

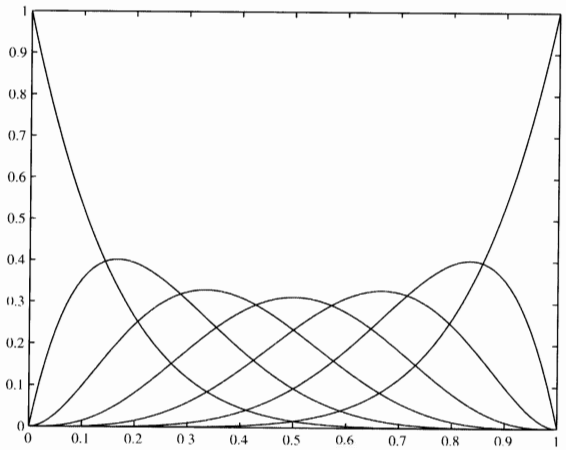


Figure 90 The seven basic Bernstein polynomials of degree six, $\binom{6}{k} x^k (1-x)^{6-k}$, $k = 0, \ldots, 6$.

$$p_n(x) = \sum_{k=0}^{n} c_k r_k(x) \qquad f(x) = \sum_{k=0}^{n} f(x) r_k(x).$$

Then we divide the sum $p_n - f = \sum (c_k - f) r_k$ into the terms where k/n is near x, and other terms where k/n is far from x. More precisely, given $\epsilon > 0$ we use uniform continuity of f on $[0, 1]$ to find $\delta > 0$ such that $|t - s| < \delta$ implies $|f(t) - f(s)| < \epsilon/2$. Then we set

$$K_1 = \{k \in \{0, \ldots, n\} : \left| \frac{k}{n} - x \right| < \delta\} \quad \text{and} \quad K_2 = \{0, \ldots n\} \setminus K_1.$$

This gives

$$\begin{aligned} |p_n(x) - f(x)| &\le \sum_{k=0}^{n} |c_k - f(x)| \, r_k(x) \\ &= \sum_{k \in K_1} |c_k - f(x)| \, r_k(x) + \sum_{k \in K_2} |c_k - f(x)| \, r_k(x). \end{aligned}$$

The factors $|c_k - f(x)|$ in the first sum are $< \epsilon/2$ since $c_k = f(k/n)$ and k/n differs from x by $< \delta$. Since the sum of all the terms r_k is 1 and the terms are non-negative, the first sum is $< \epsilon/2$. To estimate the second sum, use (3) to write

$$\begin{aligned} nx(1-x) &= \sum_{k=0}^{n} (k - nx)^2 r_k(x) \ge \sum_{k \in K_2} (k - nx)^2 r_k(x) \\ &\ge \sum_{k \in K_2} (n\delta)^2 r_k(x), \end{aligned}$$

since $k \in K_2$ implies that $(k - nx)^2 \ge (n\delta)^2$. This implies that

$$\sum_{k \in K_2} r_k(x) \le \frac{nx(1-x)}{(n\delta)^2} \le \frac{1}{4n\delta^2}$$

since $\max x(1-x) = 1/4$ as x varies in $[0, 1]$. The factors $|c_k - f(x)|$ in the second sum are at most $2M$ where $M = \|f\|$. Thus the second sum is

$$\sum_{k \in K_2} |c_k - f(x)| \, r_k(x) \le \frac{M}{2n\delta^2} \le \frac{\epsilon}{2}$$

when n is large, completing the proof that $|p_n(x) - f(x)| < \epsilon$ when n is large.

It remains to check the identities (2) and (3). The binomial coefficients satisfy

$$(4) \qquad (x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k},$$

which becomes (2) if we set $y = 1 - x$. On the other hand, if we fix y and differentiate (4) with respect to x once, and then again, we get

$$(5) \qquad n(x+y)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} k x^{k-1} y^{n-k},$$

$$(6) \qquad n(n-1)(x+y)^{n-2} = \sum_{k=0}^{n} \binom{n}{k} k(k-1) x^{k-2} y^{n-k}.$$

Note that the bottom term in (5) and the bottom two terms in (6) are 0. Multiplying (5) by x and (6) by x^2 and then setting $y = 1 - x$ in both equations gives

$$(7) \qquad nx = \sum_{k=0}^{n} \binom{n}{k} k x^k (1-x)^{n-k} = \sum_{k=0}^{n} k r_k(x),$$

$$(8) \qquad n(n-1)x^2 = \sum_{k=0}^{n} \binom{n}{k} k(k-1) x^k (1-x)^{n-k} = \sum_{k=0}^{n} k(k-1) r_k(x).$$

The last sum is $\sum k^2 r_k(x) - \sum k r_k(x)$. Hence (7), (8) become

$$(9) \qquad \sum_{k=0}^{n} k^2 r_k(x) = n(n-1)x^2 + \sum_{k=0}^{n} k r_k(x) = n(n-1)x^2 + nx.$$

Using (2), (7), (9), we get

$$\begin{aligned} &\sum_{k=0}^{n} (k - nx)^2 r_k(x) \\ &= \sum_{k=0}^{n} k^2 r_k(x) - 2nx \sum_{k=0}^{n} k r_k(x) + (nx)^2 \sum_{k=0}^{n} r_k(x) \\ &= n(n-1)x^2 + nx - 2(nx)^2 + (nx)^2 \\ &= -nx^2 + nx = nx(1-x), \end{aligned}$$

as claimed in (3). □

Proof #2 Let $f \in C^0([0,1], \mathbb{R})$ be given and let $g(x) = f(x) - (mx + b)$ where

$$m = \frac{f(1) - f(0)}{1} \quad \text{and} \quad b = f(0).$$

Then $g \in C^0$ and $g(0) = 0 = g(1)$. If we can approximate g arbitrarily well by polynomials, then the same is true of f since $mx + b$ is a polynomial. In other words it is no loss of generality to assume that $f(0) = f(1) = 0$ in the first place. Also, we extend f to all of $\mathbb{R}$ by defining $f(x) = 0$ for all $x \in \mathbb{R} \setminus [0,1]$. Then we consider a function

$$\beta_n(t) = b_n(1 - t^2)^n \qquad -1 \le t \le 1,$$

where the constant b_n is chosen so that $\int_{-1}^{1} \beta_n(t)\,dt = 1$. As shown in Figure 91, β_n is a kind of polynomial bump function. Set

$$P_n(x) = \int_{-1}^{1} f(x+t)\beta_n(t)\,dt.$$

This is a weighted average of the values of f using the weight function β_n. We claim that P_n is a polynomial and $P_n(x) \rightrightarrows f(x)$ as $n \to \infty$.

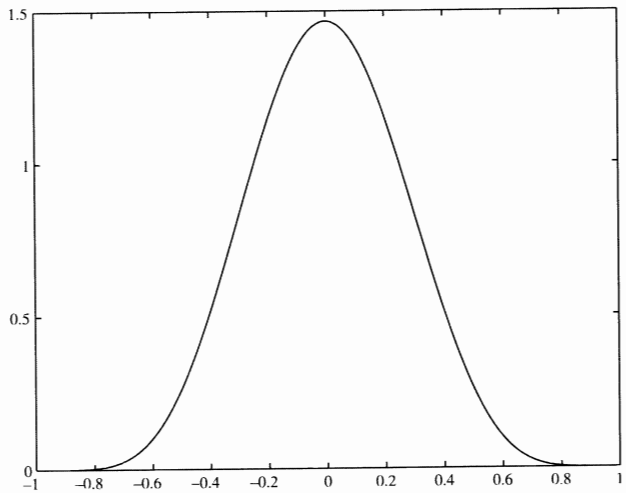


Figure 91 The graph of the function $\beta_6(t) = 1.467(1 - t^2)^6$.

To check that P_n is a polynomial we use a change of variables, $u = x + t$. Then

$$P_n(x) = \int_{x-1}^{x+1} f(u)\beta_n(u - x)\,du = \int_{0}^{1} f(u)\beta_n(x - u)\,du$$

since $f = 0$ outside of $[0,1]$. The function $\beta_n(x - u)$ is a polynomial in

x whose coefficients are polynomials in u. The powers of x pull out past the integral and we are left with these powers of x multiplied by numbers, the integrals of the polynomials in u times $f(u)$. In other words, by merely inspecting the last formula, it becomes clear that $P_n(x)$ is a polynomial in x.

To check that $P_n \rightrightarrows f$ as $n \to \infty$, we need to estimate $\beta_n(t)$. We claim that if $\delta > 0$ then

$$\beta_n(t) \rightrightarrows 0 \text{ as } n \to \infty \text{ and } \delta \le |t| \le 1. \tag{10}$$

This is clear from Figure 91. Proceeding more rigorously, we have

$$1 = \int_{-1}^{1} \beta_n(t)\,dt \ge \int_{-1/\sqrt{n}}^{1/\sqrt{n}} b_n(1-t^2)^n\,dt \ge b_n \frac{2}{\sqrt{n}} \left(1 - \frac{1}{n}\right)^n.$$

Since $1/e = \lim_{n\to\infty} (1 - 1/n)^n$, we see that for some constant c and all n,

$$b_n \le c\sqrt{n}.$$

See also Exercise 29. Hence if $\delta \le |t| \le 1$ then

$$\beta_n(t) = b_n(1 - t^2)^n \le c\sqrt{n}(1-\delta^2)^n \to 0 \text{ as } n \to \infty,$$

due to the fact that $\sqrt{n}$ tends to ∞ more slowly than $(1-\delta^2)^{-n}$ as $n \to \infty$. This proves (10).

From (10) we deduce that $P_n \rightrightarrows f$, as follows:

Let $\epsilon > 0$ be given. Uniform continuity of f gives $\delta > 0$ such that $|t| < \delta$ implies $|f(x+t) - f(x)| < \epsilon/2$. Since β_n has integral 1 on $[-1, 1]$ we have

$$\begin{aligned}
|P_n(x) - f(x)| &= \left| \int_{-1}^{1} (f(x+t) - f(x))\beta_n(t)\,dt \right| \\
&\le \int_{-1}^{1} |f(x+t) - f(x)|\,\beta_n(t)\,dt \\
&= \int_{|t|<\delta} |f(x+t) - f(x)|\,\beta_n(t)\,dt + \int_{|t|\ge\delta} |f(x+t) - f(x)|\,\beta_n(t)\,dt.
\end{aligned}$$

The first integral is $< \epsilon/2$, while the second is at most $2M \int_{|t|\ge\delta} \beta_n(t)\,dt$. By (10), the second integral is $< \epsilon/2$ when n is large. Thus $P_n \rightrightarrows f$ as claimed. □

Next we see how to extend this result to functions defined on a compact metric space M instead of merely on an interval. A subset $\mathcal{A}$ of $C^0M = C^0(M, \mathbb{R})$ is a **function algebra** if it is closed under addition, scalar multiplication, and function multiplication. That is, if $f, g \in \mathcal{A}$ and c is a constant then $f + g$, cf, and $f \cdot g$ belong to $\mathcal{A}$. For example, the set of polynomials is a function algebra. The function algebra **vanishes at a point** p if $f(p) = 0$ for all $f \in \mathcal{A}$. For example, the function algebra of all polynomials with zero constant term vanishes at $x = 0$. The function algebra **separates points** if for each pair of distinct points $p_1, p_2 \in M$ there is a function $f \in \mathcal{A}$ such that $f(p_1) \ne f(p_2)$. For example, the function algebra of all trigonometric polynomials separates points of $[0, 2\pi)$ and vanishes nowhere.

19 Stone-Weierstrass Theorem *If M is a compact metric space and $\mathcal{A}$ is a function algebra in C^0M that vanishes nowhere and separates points then $\mathcal{A}$ is dense in C^0M.*

Although the Weierstrass approximation theorem is a special case of the Stone-Weierstrass theorem, the proof of the latter does not stand on its own; it depends crucially on the former. We also need two lemmas.

20 Lemma *If $\mathcal{A}$ vanishes nowhere and separates points then, given distinct points $p_1, p_2 \in M$, and given constants c_1, c_2, there exists a function $f \in \mathcal{A}$ such that $f(p_1) = c_1$ and $f(p_2) = c_2$.*

Proof Choose $g_1, g_2 \in \mathcal{A}$ that satisfy $g_1(p_1) \ne 0 \ne g_2(p_2)$. Then $g = g_1^2 + g_2^2$ belongs to $\mathcal{A}$ and $g(p_1) \ne 0 \ne g(p_2)$. Let $h \in \mathcal{A}$ separate p_1, p_2, and consider the matrix

$$H = \begin{bmatrix} a & ab \\ c & cd \end{bmatrix} = \begin{bmatrix} g(p_1) & g(p_1)h(p_1) \\ g(p_2) & g(p_2)h(p_2) \end{bmatrix}.$$

By construction $a, c \ne 0$ and $b \ne d$. Hence $\det H = acd - abc = ac(d-b) \ne 0$, H has rank 2, and the linear equations

$$\begin{aligned} a\xi + ab\eta &= c_1 \\ c\xi + cd\eta &= c_2 \end{aligned}$$

have a solution (ξ, η). Then $f = \xi g + \eta gh$ belongs to $\mathcal{A}$ and $f(p_1) = c_1$, $f(p_2) = c_2$. □

21 Lemma *The closure of a function algebra in C^0M is a function algebra.*

Proof Clear enough. □

Proof of the Stone-Weierstrass Theorem. Let $\mathcal{A}$ be a function algebra in C^0M that vanishes nowhere and separates points. We must show that $\mathcal{A}$ is dense in C^0M: given $F \in C^0M$ and $\epsilon > 0$ we must find $G \in \mathcal{A}$ such that for all $x \in M$,

$$F(x) - \epsilon < G(x) < F(x) + \epsilon. \tag{11}$$

First we observe that

$$f \in \overline{\mathcal{A}} \quad \Rightarrow \quad |f| \in \overline{\mathcal{A}} \tag{12}$$

where $\overline{\mathcal{A}}$ denotes the closure of $\mathcal{A}$ in C^0M. Let $\epsilon > 0$ be given. According to the Weierstrass approximation theorem, there exists a polynomial $p(y)$ such that

$$\sup\{|p(y) - |y|| : |y| \leq \|f\|\} < \frac{\epsilon}{2} \tag{13}$$

After all, $|y|$ is a continuous function defined on the interval $[-\|f\|, \|f\|]$. The constant term of $p(y)$ is at most $\epsilon/2$ since $|p(0) - |0|| < \epsilon/2$. Let $q(y) = p(y) - p(0)$. Then $q(y)$ is a polynomial with zero constant term and (13) becomes

$$|q(y) - |y|| < \epsilon. \tag{14}$$

Write $q(y) = a_1 y + a_2 y^2 + \cdots + a_n y^n$ and

$$g = a_1 f + a_2 f^2 + \cdots + a_n f^n.$$

Lemma 21 states that $\overline{\mathcal{A}}$ is an algebra, so $g \in \overline{\mathcal{A}}$.† Besides, if $x \in M$ and $y = f(x)$ then

$$|g(x) - |f(x)|| = |q(y) - |y|| < \epsilon.$$

Hence $|f| \in \overline{\overline{\mathcal{A}}} = \overline{\mathcal{A}}$ as claimed in (12).

Next we observe that if f, g belong to $\overline{\mathcal{A}}$, then $\max(f, g)$ and $\min(f, g)$ also belong to $\overline{\mathcal{A}}$. For

$$\max(f, g) = \frac{f + g}{2} + \frac{|f - g|}{2}$$
$$\min(f, g) = \frac{f + g}{2} - \frac{|f - g|}{2}.$$

† Since a function algebra need not contain constant functions, it was important that q has no constant term. One should not expect that $g = a_0 + a_1 f + \cdots + a_n f^n$ belongs to $\overline{\mathcal{A}}$.

Repetition shows that the maximum and minimum of any finite number of functions in $\overline{\mathcal{A}}$ also belongs to $\overline{\mathcal{A}}$.

Now we return to (11). Let $F \in C^0M$ and $\epsilon > 0$ be given. We are trying to find $G \in \overline{\mathcal{A}}$ whose graph lies in the ϵ-tube around the graph of F. Fix any distinct points $p, q \in M$. According to Lemma 20 we can find a function in $\mathcal{A}$ with given values at p, q, say $H_{pq} \in \mathcal{A}$ satisfies

$$H_{pq}(p) = F(p) \quad \text{and} \quad H_{pq}(q) = F(q).$$

Fix p and let q vary. Each $q \in M$ has a neighborhood U_q such that

$$x \in U_q \quad \Rightarrow \quad F(x) - \epsilon < H_{pq}(x). \tag{15}$$

For $H_{pq}(x) - F(x) + \epsilon$ is a continuous function of x which is positive at $x = q$. The function H_{pq} locally **supersolves** (11). See Figure 92.

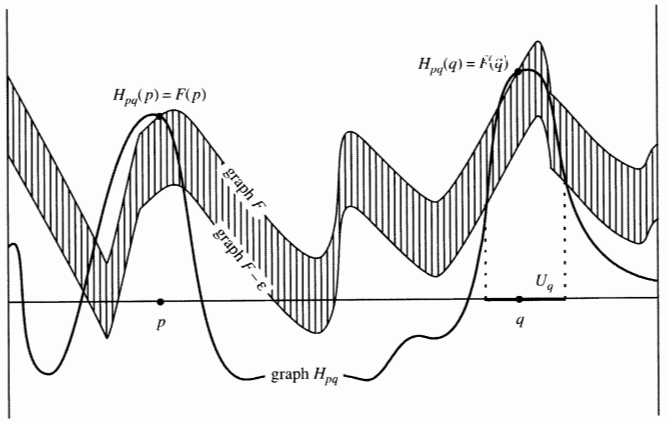


Figure 92 In a neighborhood of q, H_{pq} supersolves (11) in the sense of (15).

Compactness of M implies that finitely many of these neighborhoods U_q cover M, say $U_{q_1}, \ldots, U_{q_n}$. Define

$$G_p(x) = \max(H_{pq_1}(x), \ldots, H_{pq_n}(x)).$$

Then $G_p \in \overline{\mathcal{A}}$ and

(16) $$G_p(p) = F(p) \quad \text{and} \quad F(x) - \epsilon < G_p(x)$$

for all $x \in M$. See Figure 93.

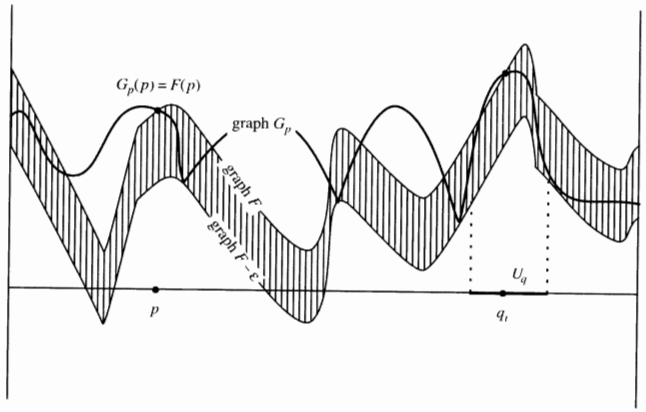


Figure 93 G_p is the maximum of H_{pq_i}, $i = 1, \ldots, n$.

Continuity implies that each p has a neighborhood V_p such that

(17) $$x \in V_p \quad \Rightarrow \quad G_p(x) < F(x) + \epsilon.$$

See Figure 94. By compactness, finitely many of these neighborhoods cover

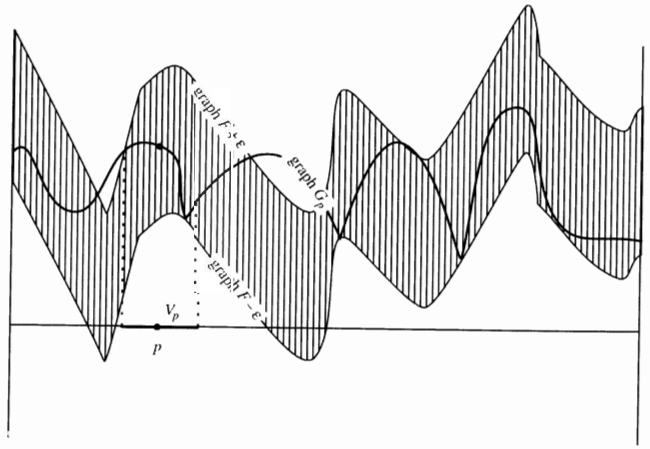


Figure 94 $G_p(p) = F(p)$ and G_p supersolves (11) everywhere.

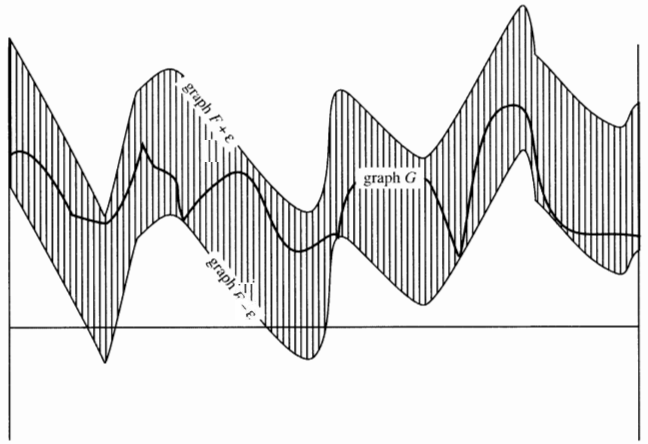


Figure 95 The graph of G lies in the ϵ-tube around the graph of F.

M, say $V_{p_1}, \ldots, V_{p_m}$. Set

$$G(x) = \min(G_{p_1}(x), \ldots, G_{p_m}(x)).$$

We know that $G \in \overline{\mathcal{A}}$ and (16), (17) imply (11). See Figure 95. □

22 Corollary *Any 2π-periodic continuous function of $x \in \mathbb{R}$ can be uniformly approximated by a* ***trigonometric polynomial***

$$T(x) = \sum_{k=0}^{n} a_k \cos kx + \sum_{k=0}^{n} b_k \sin kx.$$

Proof Think of $[0, 2\pi)$ parameterizing the circle S^1 by $x \mapsto (\cos x, \sin x)$. The circle is compact, and 2π-periodic continuous functions on $\mathbb{R}$ become continuous functions on S^1. The trigonometric polynomials on S^1 form an algebra $\mathcal{T} \subset C^0 S^1$ that vanishes nowhere and separates points. The Stone-Weierstrass theorem implies that $\mathcal{T}$ is dense in $C^0 S^1$. □

Here is a typical application of the Stone-Weierstrass Theorem: Consider a continuous vector field $F : \Delta \to \mathbb{R}^2$ where Δ is the closed unit disc in the plane, and suppose that we want to approximate F by a vector field that vanishes (equals zero) at most finitely often. A simple way to do so is to approximate F by a polynomial vector field G. Real polynomials in two variables are finite sums

$$P(x, y) = \sum_{i,j=0}^{n} c_{ij} x^i y^j$$

where the c_{ij} are constants. They form a function algebra $\mathcal{A}$ in $C^0(\Delta, \mathbb{R})$ that

separates points and vanishes nowhere. By the Stone-Weierstrass Theorem, $\mathcal{A}$ is dense in C^0, so we can approximate the components of $F = (F_1, F_2)$ by polynomials

$$F_1 \doteq P \quad F_2 \doteq Q.$$

The vector field (P, Q) then approximates F. Changing the coefficients of P by a small amount ensures that P and Q have no common polynomial factor and F vanishes at most finitely often.

5 Contractions and ODE's

Fixed point theorems are of great use in the applications of analysis, including the basic theory of vector calculus such as the general implicit function theorem. If $f : M \to M$ and for some $p \in M$, $f(p) = p$, then p is a **fixed point** of f. When must f have a fixed point? This question has many answers, and the two most famous are given in the next two theorems.

Let M be a metric space. A **contraction** of M is a mapping $f : M \to M$ such that for some constant $k < 1$ and all $x, y \in M$,

$$d(f(x), f(y)) \leq kd(x, y).$$

23 Banach Contraction Principle *Suppose that $f : M \to M$ is a contraction and the metric space M is complete. Then f has a unique fixed point p and for any $x \in M$, $f^n(x) = f \circ f \circ \cdots \circ f(x) \to p$ as $n \to \infty$.*

Brouwer Fixed Point Theorem *Suppose that $f : B^m \to B^m$ is continuous where B^m is the closed unit ball in $\mathbb{R}^m$. Then f has a fixed point $p \in B^m$.*

The proof of the first result is easy, the second not. See Figure 96 to picture a contraction.

Proof #1 of the Banach Contraction Principle Beautiful, simple, and dynamical! See Figure 96. Choose any $x_0 \in M$ and define $x_n = f^n(x_0)$. We claim that for all $n \in \mathbb{N}$

$$d(x_n, x_{n+1}) \leq k^n d(x_0, x_1). \tag{18}$$

This is easy:

$$\begin{aligned} d(x_n, x_{n+1}) &\leq kd(f(x_{n-1}), f(x_n)) \leq k^2 d(f(x_{n-2}), f(x_{n-1})) \\ &\leq \cdots \leq k^n d(x_0, x_1). \end{aligned}$$

From this and a geometric series type of estimate, it follows that the sequence (x_n) is Cauchy. For let $\epsilon > 0$ be given. Choose N large enough

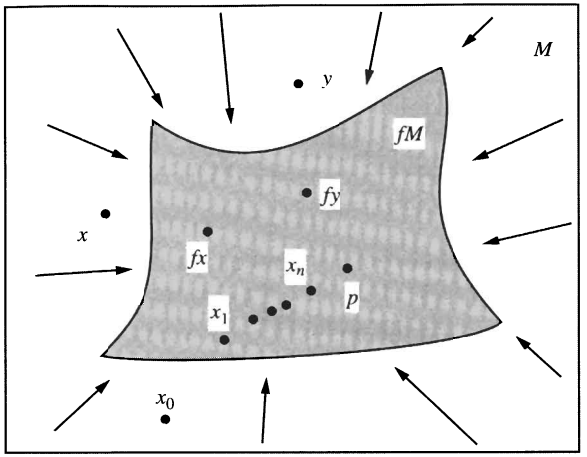


Figure 96 f contracts M toward the fixed point p.

that

$$\frac{k^N}{1-k} d(x_0, x_1) < \epsilon. \tag{19}$$

Note that (19) needs the hypothesis $k < 1$. If $N \leq m \leq n$ then

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \cdots + d(x_{n-1}, x_n) \\ &\leq k^m d(x_0, x_1) + k^{m+1} d(x_0, x_1) + \cdots + k^{n-1} d(x_0, x_1) \\ &\leq k^m (1 + k + \cdots + k^{n-m-1}) d(x_0, x_1) \\ &\leq k^N \Big(\sum_{\ell=0}^{\infty} k^\ell \Big) d(x_0, x_1) = \frac{k^N}{1-k} d(x_0, x_1) < \epsilon. \end{aligned}$$

Thus (x_n) is Cauchy. Since M is complete, x_n converges to some $p \in M$ as $n \to \infty$. Then

$$\begin{aligned} d(p, f(p)) &\leq d(p, x_n) + d(x_n, f(x_n)) + d(f(x_n), f(p)) \\ &\leq d(p, x_n) + k^n d(x_0, x_1) + kd(x_n, p) \to 0 \text{ as } n \to \infty. \end{aligned}$$

Since $d(p, f(p))$ is independent of n, $d(p, f(p)) = 0$ and $p = f(p)$. This proves the existence of the fixed point. Uniqueness is immediate. After all, how can two points simultaneously stay fixed and move closer together? □

Proof #2 of the Banach Contraction Principle — sketch Choose any point $x_0 \in M$ and choose r_0 so large that $f(\overline{M_{r_0}(x_0)}) \subset M_{r_0}(x_0)$. Let $B_0 = \overline{M_{r_0}(x_0)}$ and $B_n = \overline{f^n(B_{n-1})}$. The diameter of B_n is at most $k^n \operatorname{diam}(B_0)$, and this tends to 0 as $n \to \infty$. The sets B_n nest downward as $n \to \infty$ and f sends B_n inside B_{n+1}. Since M is complete, this implies that $\bigcap B_n$ is a single point, say p, and $f(p) = p$. □

Proof of Brouwer's Theorem in dimension one The closed unit 1-ball is the interval $[-1, 1]$ in $\mathbb{R}$. If $f : [-1, 1] \to [-1, 1]$ is continuous then so is $g(x) = x - f(x)$. At the endpoints ± 1, we have $g(-1) \leq 0 \leq g(1)$. By the intermediate value theorem, there is a point $p \in [-1, 1]$ such that $g(p) = 0$. That is, $f(p) = p$. □

The proof in higher dimensions is harder. One proof is a consequence of the general Stokes' Theorem, and is given in Chapter 5. Another depends on algebraic topology, a third on differential topology.

Ordinary Differential Equations

The qualitative theory of ordinary differential equations (ODE's) begins with the basic existence/uniqueness theorem in ODE's, Picard's Theorem. Throughout, U is an open subset of m-dimensional Euclidean space $\mathbb{R}^m$. In geometric terms, an ODE is a vector field F defined on U. We seek a **trajectory** γ of F through a given $p \in U$, i.e., $\gamma : (a, b) \to U$ is differentiable and solves the ODE with **initial condition** p,

$$\gamma'(t) = F(\gamma(t)) \text{ and } \gamma(0) = p. \tag{20}$$

See Figure 97. In this notation we think of the vector field F defining at each $x \in U$ a vector $F(x)$ whose foot lies at x and to which γ must be tangent. The vector $\gamma'(t)$ is $(\gamma_1'(t), \ldots, \gamma_m'(t))$ where $\gamma_1, \ldots, \gamma_m$ are the components of γ. The trajectory $\gamma(t)$ describes how a particle travels with prescribed velocity F. At each time t, $\gamma(t)$ is the position of the particle; its velocity there is exactly the vector F at that point. Intuitively, trajectories should exist because particles do move.

The contraction principle gives a way to find trajectories of vector fields, or — what is the same thing to solve ODE's. We will assume that F satisfies a **Lipschitz condition** — there is a constant L such that for all points $x, y \in U$

$$|F(x) - F(y)| \leq L\,|x - y|\,.$$

Here, $|\ \ |$ refers to the Euclidean length of a vector. F, x, y are all vectors

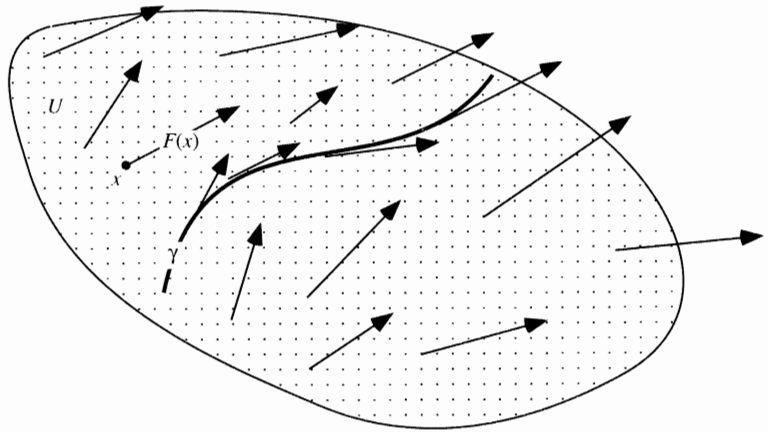


Figure 97 γ is always tangent to the vector field F.

in $\mathbb{R}^m$. It follows that F is continuous. The Lipschitz condition is stronger than continuity, but still fairly mild. Any differentiable vector field with a bounded derivative is Lipschitz.

24 Picard's Theorem *Given $p \in U$ there exists an F-trajectory $\gamma(t)$ in U through p. This means that $\gamma : (a, b) \to U$ solves (20). Locally, γ is unique.*

To prove Picard's Theorem it is convenient to re-express (20) as an integral equation and to do this we make a brief digression about vector-valued integrals. Let's recall four key facts about integrals of real valued functions of a real variable, $y = f(x)$, $a \leq x \leq b$.

(a) $\int_a^b f(x)\,dx$ is approximated by Riemann sums $R = \sum f(t_k)\Delta x_k$.
(b) Continuous functions are integrable.
(c) If $f'(x)$ exists and is continuous then $\int_a^b f'(x)\,dx = f(b) - f(a)$.
(d) $\left|\int_a^b f(x)\,dx\right| \leq M(b - a)$ where $M = \sup|f(x)|$.

The Riemann sum R in (a) has $a = x_0 \leq \cdots \leq x_{k-1} \leq t_k \leq x_k \leq \cdots \leq x_n = b$ and all the $\Delta x_k = x_k - x_{k-1}$ are small.

Given a vector-valued function of a real variable

$$f(x) = (f_1(x), \ldots, f_m(x)),$$

$a \leq x \leq b$, we define its integral componentwise as the vector of integrals

$$\int_a^b f(x)\,dx = \left(\int_a^b f_1(x)\,dx, \ldots, \int_a^b f_m(x)\,dx\right).$$

Corresponding to (a) - (d) we have

(a′) $\int_a^b f(x)\,dx$ is approximated by $R = (R_1, \ldots, R_m)$, with R_j a Riemann sum for f_j.

(b′) Continuous vector-valued functions are integrable.

(c′) If $f'(x)$ exists and is continuous, then $\int_a^b f'(x)\,dx = f(b) - f(a)$.

(d′) $\left|\int_a^b f(x)\,dx\right| \leq M(b-a)$ where $M = \sup|f(x)|$.

(a′), (b′), (c′) are clear enough. To check (d′) we write

$$\begin{aligned} R &= \sum R_j e_j = \sum_j \sum_k f_j(t_k)\Delta x_k e_j \\ &= \sum_k \sum_j f_j(t_k) e_j \Delta x_k = \sum_k f(t_k)\Delta x_k \end{aligned}$$

where $e_1, \ldots, e_m$ is the standard vector basis for $\mathbb{R}^m$. Thus,

$$|R| \leq \sum_k |f(t_k)|\,\Delta x_k \leq \sum_k M\Delta x_k = M(b-a),$$

By (a′), R approximates the integral, which implies (d′). (Note that a weaker inequality with M replaced by $\sqrt{m}M$ follows immediately from (d). This weaker inequality would suffice for most of what we do — but it is inelegant.)

Now consider the following integral version of (20),

$$\gamma(t) = p + \int_0^t F(\gamma(s))\,ds. \tag{21}$$

A solution of (21) is by definition any continuous curve $\gamma : (a, b) \to U$ for which (21) holds identically in $t \in (a, b)$. By (b′) any solution of (21) is automatically differentiable and its derivative is $F(\gamma(t))$. That is, any solution of (21) solves (20). The converse is also clear, so solving (20) is equivalent to solving (21).

Proof of Picard's Theorem Since F is continuous, there exists a compact neighborhood $N = \overline{N}_r(p)$ and a constant M such that $|F(x)| \leq M$ for all $x \in N$. Choose $\tau > 0$ such that

$$\tau M \leq r \text{ and } \tau L < 1. \tag{22}$$

Consider the set $\mathscr{C}$ of all continuous functions $\gamma : [-\tau, \tau] \to N$. With respect to the metric

$$d(\gamma, \sigma) = \sup\{|\gamma(t) - \sigma(t)| : t \in [-\tau, \tau]\}$$

the set $\mathscr{C}$ is a complete metric space. Given $\gamma \in \mathscr{C}$, define

$$\Phi(\gamma)(t) = p + \int_0^t F(\gamma(s))\,ds.$$

Solving (21) is the same as finding γ such that $\Phi(\gamma) = \gamma$. That is, we seek a fixed point of Φ.

We just need to show that Φ is a contraction of $\mathscr{C}$. Does Φ send $\mathscr{C}$ into itself? Given $\gamma \in \mathscr{C}$ we see that $\Phi(\gamma)(t)$ is a continuous (in fact differentiable) vector-valued function of t and that by (22),

$$|\Phi(\gamma)(t) - p| = \left|\int_0^t F(\gamma(s))\,ds\right| \leq \tau M \leq r\,.$$

Therefore, Φ does send $\mathscr{C}$ into itself.

Φ contracts $\mathscr{C}$ because

$$\begin{aligned} d(\Phi(\gamma), \Phi(\sigma)) &= \sup_t \left|\int_0^t F(\gamma(s)) - F(\sigma(s))\,ds\right| \\ &\leq \tau \sup_s |F(\gamma(s)) - F(\sigma(s))| \\ &\leq \tau \sup_s L\,|\gamma(s) - \sigma(s)| \leq \tau L d(\gamma, \sigma) \end{aligned}$$

and $\tau L < 1$ by (22). Therefore Φ has a fixed point γ, and $\Phi(\gamma) = \gamma$ implies that $\gamma(t)$ solves (21), which implies that γ is differentiable and solves (20).

Any other solution $\sigma(t)$ of (20) defined on the interval $[-\tau, \tau]$ also solves (21) and is a fixed point of Φ, $\Phi(\sigma) = \sigma$. Since a contraction mapping has a unique fixed point, $\gamma = \sigma$, which is what local uniqueness means. □

The F-trajectories define a **flow** in the following way: To avoid the possibility that trajectories cross the boundary of U (they "escape from U") or become unbounded in finite time (they "escape to infinity") we assume that U is all of $\mathbb{R}^m$. Then trajectories can be defined for all time $t \in \mathbb{R}$. Let $\gamma(t, p)$ denote the trajectory through p. Imagine all points $p \in \mathbb{R}^m$ moving *in unison* along their trajectories as t increases. They are leaves on a river, motes in a breeze. The point $p_1 = \gamma(t_1, p)$ at which p arrives after time t_1 moves according to $\gamma(t, p_1)$. Before p arrives at p_1, however, p_1 has already gone elsewhere. This is expressed by the flow equation

$$\gamma(t, p_1) = \gamma(t + t_1, p).$$

See Figure 98.

The flow equation is true because as functions of t both sides of the equation are F-trajectories through p_1, and the F-trajectory through a point

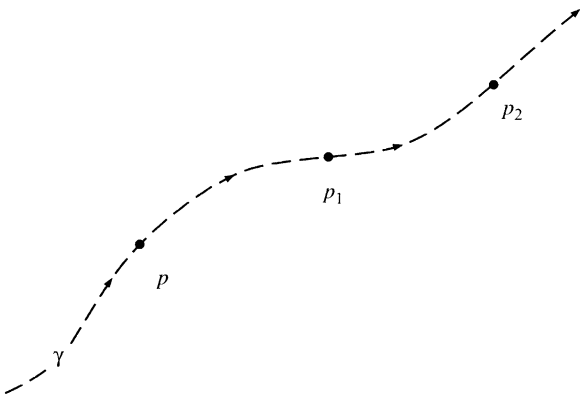


Figure 98 The time needed to flow from from p to p_2 is the sum of the times needed to flow from p to p_1 and from p_1 to p_2.

is locally unique. It is revealing to rewrite the flow equation with different notation. Setting $\varphi_t(p) = \gamma(t, p)$ gives

$$\varphi_{t+s}(p) = \varphi_t(\varphi_s(p)) \text{ for all } t, s \in \mathbb{R}.$$

φ_t is called the ***t*-advance map**. It specifies where each point moves after time t. See Figure 99. The flow equation states that $t \mapsto \varphi_t$ is a group

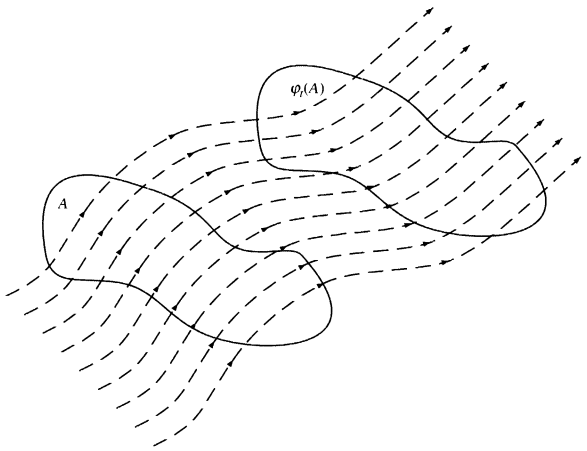


Figure 99 The t-advance map shows how a set A flows to a set $\varphi_t(A)$.

homomorphism from $\mathbb{R}$ into the group of motions of $\mathbb{R}^m$. In fact each φ_t is a homeomorphism of $\mathbb{R}^m$ onto itself and its inverse is φ_{-t}. For $\varphi_{-t} \circ \varphi_t = \varphi_0$ and φ_0 is the time-zero map where nothing moves at all, $\varphi_0 =$ identity map.

6* Analytic Functions

Recall from Chapter 3 that a function $f : (a, b) \to \mathbb{R}$ is **analytic** if it can be expressed locally as a power series. For each $x \in (a, b)$ there exists a convergent power series $\sum c_k h^k$ such that for all $x + h$ near x,

$$f(x + h) = \sum_{k=0}^{\infty} c_k h^k.$$

As we have shown previously, every analytic function is smooth but not every smooth function is analytic. In this section we give a necessary and sufficient condition that a smooth function be analytic. It involves the speed with which the r^{th} derivative grows as $r \to \infty$.

Let $f : (a, b) \to \mathbb{R}$ be smooth. The **Taylor series** for f at $x \in (a, b)$ is

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(x)}{k!} h^k.$$

Let $I = [x - \sigma, x + \sigma]$ be a subinterval of (a, b), $\sigma > 0$, and denote by M_r the maximum of $\left|f^{(r)}(t)\right|$ for $t \in I$. The **derivative growth rate** of f on I is

$$\alpha = \limsup_{r \to \infty} \sqrt[r]{\frac{M_r}{r!}}.$$

Clearly, $\sqrt[r]{\left|f^{(r)}(x)\right|/r!} \le \sqrt[r]{M_r/r!}$, so the radius of convergence

$$R = \frac{1}{\limsup_{r \to \infty} \sqrt[r]{\dfrac{\left|f^{(r)}(x)\right|}{r!}}}$$

of the Taylor series at x satisfies

$$\frac{1}{\alpha} \le R.$$

In particular, if α is finite the radius of convergence of the Taylor series is positive.

25 Theorem *If $\alpha\sigma < 1$, then the Taylor series converges uniformly to f on the interval I.*

Proof Choose $\delta > 0$ such that $(\alpha+\delta)\sigma < 1$. The Taylor remainder formula from Chapter 3, applied to the $(r-1)^{\text{st}}$ order remainder, gives

$$f(x+h) - \sum_{k=0}^{r-1} \frac{f^{(k)}(x)}{k!} h^k = \frac{f^{(r)}(\theta)}{r!} h^r$$

for some θ between x and $x+h$. Thus, for r large

$$\left| f(x+h) - \sum_{k=0}^{r-1} \frac{f^{(k)}(x)}{k!} h^k \right| \le \frac{M_r}{r!}\sigma^r = \left(\left(\frac{M_r}{r!}\right)^{1/r} \sigma \right)^r \le ((\alpha+\delta)\sigma)^r.$$

Since $(\alpha+\delta)\sigma < 1$, the Taylor series converges uniformly to $f(x+h)$ on I. □

26 Theorem *If f is expressed as a convergent power series $f(x+h) = \sum c_k h^k$ with radius of convergence $R > \sigma$, then f has bounded derivative growth rate on I.*

The proof of Theorem 26 uses two estimates about the growth rate of factorials. If you know Stirling's formula they are easy, but we prove them directly.

(23) $$\lim_{r\to\infty} \sqrt[r]{\frac{r^r}{r!}} = e$$

(24) $$0 < \lambda < 1 \quad \Rightarrow \quad \limsup_{r\to\infty} \sqrt[r]{\sum_{k=r}^{\infty} \binom{k}{r} \lambda^k} \; < \; \infty.$$

Taking logarithms, applying the integral test, and ignoring terms that tend to zero as $r \to \infty$ gives

$$\begin{aligned} \frac{1}{r}(\log r^r - \log r!) &= \log r - \frac{1}{r}(\log r + \log(r-1) + \cdots + \log 1) \\ \sim \log r - \frac{1}{r}\int_1^r \log x\, dx &= \log r - \frac{1}{r}(x\log x - x)\Big|_1^r \\ &= 1 - \frac{1}{r}, \end{aligned}$$

which tends to 1 as $r \to \infty$. This proves (23).

To prove (24) we write $\lambda = e^{-\mu}$ for $\mu > 0$, and reason similarly:

$$\begin{aligned} \sum_{k=r}^{\infty} \binom{k}{r} \lambda^k &= \sum_{k=r}^{\infty} \frac{k(k-1)(k-2)\dots(k-r+1)}{r!} e^{-k\mu} \\ &\le \frac{1}{r!} \sum_{k=r}^{\infty} k^r e^{-k\mu} \quad \sim \quad \frac{1}{r!}\int_r^{\infty} x^r e^{-\mu x}\, dx \\ &= \frac{-1}{r!} e^{-\mu x}\left(\frac{x^r}{\mu} + \frac{r x^{r-1}}{\mu^2} + \frac{r(r-1)x^{r-2}}{\mu^3} + \cdots + \frac{r!}{\mu^{r+1}} \right)\Big|_r^{\infty} \\ &\le \frac{1}{r!} e^{-\mu r}(r+1) r^r \left(\frac{1}{\min(1,\mu)} \right)^{r+1}. \end{aligned}$$

According to (23) the r^{th} root of this quantity tends to $e^{1-\mu}/\min(1,\mu)$ as $r \to \infty$, completing the proof of (24).

Proof of Theorem 26 By assumption the power series $\sum c_k h^k$ has radius of convergence R and $\sigma < R$. Since $1/R$ is the lim sup of $\sqrt[k]{|c_k|}$ as $k \to \infty$, there is a number $\lambda < 1$ such that for all large k, $|c_k \sigma^k| \le \lambda^k$. Differentiating the series term by term with $|h| \le \sigma$ gives

$$\begin{aligned} \left| f^{(r)}(x+h) \right| &\le \sum_{k=r}^{\infty} k(k-1)(k-2)\dots(k-r+1) \left| c_k h^{k-r} \right| \\ &\le \frac{r!}{\sigma^r} \sum_{k=r}^{\infty} \binom{k}{r} \left| c_k \sigma^k \right| \le \frac{r!}{\sigma^r} \sum_{k=r}^{\infty} \binom{k}{r} \lambda^k \end{aligned}$$

for r large. Thus,

$$M_r = \sup_{|h|\le\sigma} \left| f^{(r)}(x+h) \right| \le \frac{r!}{\sigma^r} \sum_{k=r}^{\infty} \binom{k}{r} \lambda^k.$$

According to (24),

$$\alpha = \limsup_{r\to\infty} \sqrt[r]{\frac{M_r}{r!}} \le \frac{1}{\sigma} \limsup_{r\to\infty} \sqrt[r]{\sum_{k=r}^{\infty} \binom{k}{r} \lambda^k} < \infty,$$

and f has bounded derivative growth rate on I. □

From Theorems 25 and 26 we deduce the main result of this section.

27 Analyticity Theorem *A smooth function is analytic if and only if it has locally bounded derivative growth rate.*

Proof Assume that $f : (a, b) \to \mathbb{R}$ is smooth and has locally bounded derivative growth rate. Then $x \in (a, b)$ has a neighborhood N on which the derivative growth rate α is finite. Choose $\sigma > 0$ such that $I = [x - \sigma, x + \sigma] \subset N$ and $\alpha\sigma < 1$. We infer from Theorem 25 that the Taylor series for f at x converges uniformly to f on I. Hence f is analytic.

Conversely, assume that f is analytic and let $x \in (a, b)$ be given. There is a power series $\sum c_k h^k$ that converges to $f(x + h)$ for all h in some interval $(-R, R)$, $R > 0$. Choose σ, $0 < \sigma < R$. We infer from Theorem 26 that f has bounded derivative growth rate on I. □

28 Corollary *A smooth function is analytic if its derivatives are uniformly bounded.*

An example of such a function is $f(x) = \sin x$.

Proof If $\left|f^{(r)}(\theta)\right| \leq M$ for all r and θ then the derivative growth rate of f is bounded, in fact $\alpha = 0$ and $R = \infty$. □

29 Taylor's Theorem *If $f(x) = \sum c_k x^k$ and the power series has radius of convergence R, then f is analytic on $(-R, R)$.*

Proof The function f is smooth, and by Theorem 26 it has bounded derivative growth rate on $I \subset (-R, R)$. Hence it is analytic. □

Taylor's Theorem states that not only can f be expanded as a convergent power series at $x = 0$, but also at any other point $x_0 \in (-R, R)$. Other proofs of Taylor's theorem rely more heavily on series manipulations and Mertens' theorem.

The concept of analyticity extends immediately to complex functions. A function $f : D \to \mathbb{C}$ is **complex analytic** if D is an open subset of $\mathbb{C}$ and for each $z \in D$ there is a power series

$$\sum c_k \zeta^k$$

such that for all $z + \zeta$ near z,

$$f(z + \zeta) = \sum_{k=0}^{\infty} c_k \zeta^k.$$

The coefficients c_k are complex and so is the variable ζ. Convergence occurs on a disc of radius R. This lets us define e^z, $\log z$, $\sin z$, $\cos z$ for the complex

number z by setting

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!} \qquad \log(1 + z) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1} z^k}{k} \text{ when } |z| < 1$$

$$\sin z = \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k+1}}{(2k + 1)!} \qquad \cos z = \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k}}{(2k)!}.$$

It is enlightening and reassuring to derive formulas such as

$$e^{i\theta} = \cos\theta + i \sin\theta$$

directly from these definitions. (Just plug in $z = i\theta$ and use the equations $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, etc.) A key formula to check is $e^{z+w} = e^z e^w$. One proof involves a manipulation of product series, a second merely uses analyticity. Another formula is $\log(e^z) = z$.

There are many natural results about real analytic functions that can be proved by direct power series means; e.g., the sum, product, reciprocal, composite, and inverse function of analytic functions are analytic. Direct proofs, like those for the Analyticity Theorem above, involve major series manipulations. The use of complex variables leads to greatly simplified proofs of these real variable theorems, thanks to the following fact.

Real analyticity propagates to complex analyticity and complex analyticity is equivalent to complex differentiability.[†]

For it is relatively easy to check that the composition, etc., of complex differentiable functions is complex differentiable.

The analyticity concept extends even beyond $\mathbb{C}$. You may already have seen such an extension when you studied the vector linear ODE

$$x' = Ax$$

in calculus. A is a given $m \times m$ matrix and the unknown solution $x = x(t)$ is a vector function of t, on which an initial condition $x(0) = x_0$ is usually imposed. A vector ODE is equivalent to m coupled, scalar, linear ODE's. The solution $x(t)$ can be expressed as

$$x(t) = e^{tA} x_0$$

[†] A function $f : D \to \mathbb{C}$ is **complex differentiable** or **holomorphic** if D is an open subset of $\mathbb{C}$ and for each $z \in D$, the limit of

$$\frac{\Delta f}{\Delta z} = \frac{f(z + \Delta z) - f(z)}{\Delta z}$$

exists as $\Delta z \to 0$ in $\mathbb{C}$. The limit, if it exists, is a complex number.

where

$$e^{tA} = \lim_{n\to\infty} (I + tA + \frac{1}{2!}(tA)^2 + \cdots + \frac{1}{n!}(tA)^n) = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k.$$

I is the $m \times m$ identity matrix. View this series as a power series with k^{th} coefficient $t^k/k!$ and variable A. (A is a matrix variable!) The limit exists in the space of all $m \times m$ matrices, and its product with the constant vector x_0 does indeed give a vector function of t that solves the original linear ODE.

The previous series defines the exponential of a matrix as $e^A = \sum A^k/k!$. You might ask yourself — is there such a thing as the logarithm of a matrix? A function that assigns to a matrix its matrix logarithm? A power series that expresses the matrix logarithm? What about other analytic functions? Is there such a thing as the sine of a matrix? What about inverting a matrix? Is there a power series that expresses matrix inversion? Are formulas such as $\log A^2 = 2 \log A$ true? These questions are explored in nonlinear functional analysis.

A terminological point on which to insist is that the word "analytic" be defined as "locally power series expressible." In the complex case, some mathematicians define complex analyticity as complex differentiability, and although complex differentiability turns out to be equivalent to local expressibility as a complex power series, this is a very special feature of $\mathbb{C}$. In fact it is responsible for every distinction between real and complex analysis. For cross, theory consistency, then, one should use the word "analytic" to mean local power series expressible, and use "differentiable" to mean differentiable. Why confound the two ideas?

7* Nowhere Differentiable Continuous Functions

Although many continuous functions, such as $|x|$, $\sqrt[3]{x}$, and $x \sin(1/x)$ fail to be differentiable at a few points, it is quite surprising that there can exist a function which is everywhere continuous but nowhere differentiable.

30 Theorem *There exists a continuous function $f : \mathbb{R} \to \mathbb{R}$ that has a derivative at no point whatsoever.*

Proof The construction is due to Weierstrass. The letters k, m, n denote integers. Start with a **sawtooth function** $\sigma_0 : \mathbb{R} \to \mathbb{R}$ defined as

$$\sigma_0(x) = \begin{cases} x - 2n & \text{if } 2n \le x \le 2n+1 \\ 2n + 2 - x & \text{if } 2n+1 \le x \le 2n+2. \end{cases}$$

σ_0 is periodic with period 2; if $t = x + 2m$ then $\sigma_0(t) = \sigma_0(x)$. The compressed sawtooth function

$$\sigma_k(x) = \left(\frac{3}{4}\right)^k \sigma_0(4^k x)$$

has period $\pi_k = 2/4^k$. If $t = x + m\pi_k$ then $\sigma_k(t) = \sigma_k(x)$. See Figure 100.

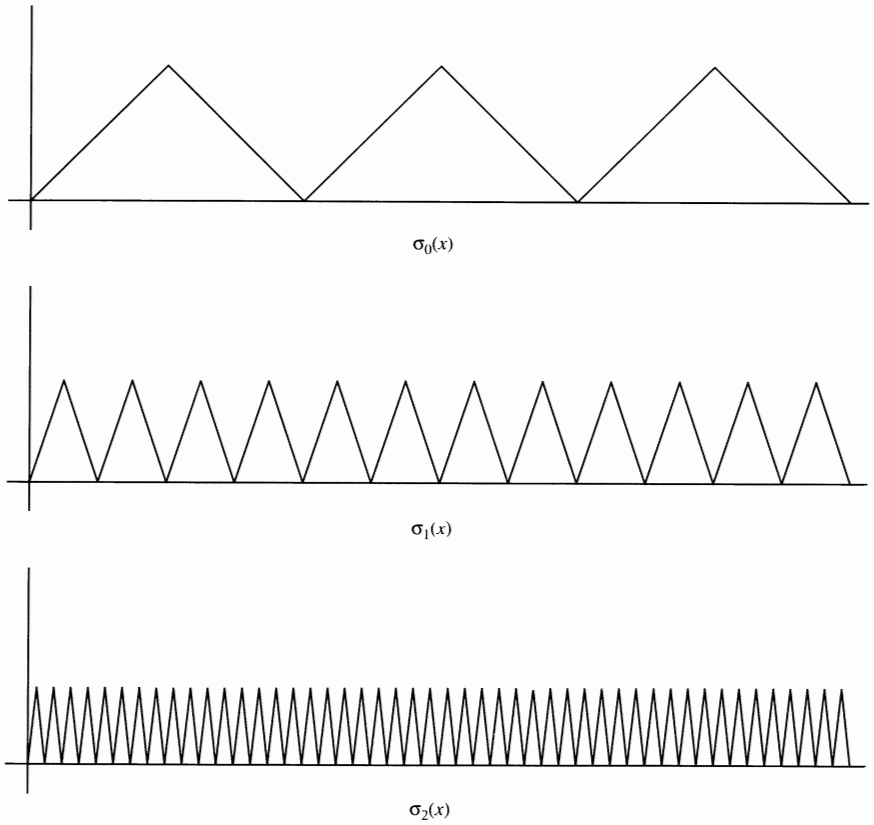


Figure 100 The graphs of the sawtooth function and two compressed sawtooth functions.

According to the M-test, the series $\sum \sigma_k(x)$ converges uniformly to a limit f, and

$$f(x) = \sum_{k=0}^{\infty} \sigma_k(x)$$

is continuous. We claim that f is nowhere differentiable. Fix an arbitrary point x, and set $\delta_n = 1/2 \cdot 4^n$. We will show that

$$\frac{\Delta f}{\Delta x} = \frac{f(x \pm \delta_n) - f(x)}{\delta_n}$$

does not converge to a limit as $\delta_n \to 0$, and thus that $f'(x)$ does not exist. The quotient is

$$\frac{\Delta f}{\Delta x} = \sum_{k=0}^{\infty} \frac{\sigma_k(x \pm \delta_n) - \sigma_k(x)}{\delta_n}.$$

There are three types of term in the series, $k > n$, $k = n$, and $k < n$. If $k > n$ then $\sigma_k(x \pm \delta_n) - \sigma_k(x) = 0$ — for δ_n is an integer multiple of the period of σ_k,

$$\delta_n = \frac{1}{2 \cdot 4^n} = 4^{k-(n+1)} \cdot \frac{2}{4^k} = 4^{k-(n+1)} \cdot \pi_k.$$

Thus the infinite series expression for $\Delta f/\Delta x$ reduces to a sum of $n + 1$ terms

$$\frac{\Delta f}{\Delta x} = \frac{\sigma_n(x \pm \delta_n) - \sigma_n(x)}{\delta_n} + \sum_{k=0}^{n-1} \frac{\sigma_k(x \pm \delta_k) - \sigma_k(x)}{\delta_k}.$$

The function σ_n is monotone on either $[x - \delta_n, x]$ or $[x, x + \delta_n]$, since it is monotone on intervals of length 4^{-n} and the contiguous interval $[x - \delta_n, x, x + \delta_n]$ at x is of length 4^{-n}. The slope of σ_n is $\pm 3^n$. Thus, either

$$\left|\frac{\sigma_n(x + \delta_n) - \sigma_n(x)}{\delta_n}\right| = 3^n \qquad \text{or} \qquad \left|\frac{\sigma_n(x - \delta_n) - \sigma_n(x)}{\delta_n}\right| = 3^n.$$

The terms with $k < n$ are crudely estimated from the slope of σ_k being $\pm 3^k$:

$$\left|\frac{\sigma_k(x \pm \delta_k) - \sigma_k(x)}{\delta_k}\right| \leq 3^k.$$

Thus

$$\left|\frac{\Delta f}{\Delta x}\right| \geq 3^n - (3^{n-1} + \cdots + 1) = 3^n - \frac{3^n - 1}{3 - 1} = \frac{1}{2}(3^n + 1),$$

which tends to ∞ as $\delta_n \to 0$, so $f'(x)$ does not exist. □

Weierstrass showed that a nowhere differentiable continuous function exists by simply writing a formula for it. Yet more amazing is the fact that *most* continuous functions (in a reasonable sense defined below) are

nowhere differentiable. If you could pick a continuous function at random, it would be nowhere differentiable.

Recall that the set $D \subset M$ is dense in M if D meets every non-empty open subset W of M, $D \cap W \neq \emptyset$. The intersection of two dense sets need not be dense; it can be empty, as is the case with $\mathbb{Q}$ and $\mathbb{Q}^c$ in $\mathbb{R}$. On the other hand if U, V are open dense sets in M then $U \cap V$ is open dense in M. For if W is any non-empty open subset of M then $U \cap W$ is a non-empty open subset of M, and by denseness of V, V meets $U \cap W$; i.e., $U \cap V \cap W$ is non-empty and $U \cap V$ meets W.

Moral Open dense sets do a good job of being dense.

The countable intersection $G = \bigcap G_n$ of open dense sets is called a **thick** (or **residual**†) subset of M, due to the following result, which we will apply in the complete metric space $C^0([a, b], \mathbb{R})$. Extending our vocabulary in a natural way we say that the complement of a thick set is **thin** (or **meager**). A subset H of M is thin if and only if it is a countable union of nowhere dense closed sets, $H = \bigcup H_n$. Clearly, thickness and thinness are topological properties. A thin set is the topological analog of a zero set (a set whose outer measure is zero).

31 Baire's Theorem *Every thick subset of a complete metric space M is dense in M. A non-empty, complete metric space is not thin: if M is the union of countably many closed sets, at least one has non-empty interior.*

If all points in a thick subset of M satisfy some condition then the condition is said to be **generic**, we also say that **most** points of M obey the condition. As a consequence of Baire's theorem and a modification of Weierstrass' construction we will prove

32 Theorem *The generic $f \in C^0 = C^0([a, b], \mathbb{R})$ is differentiable at no point of $[a, b]$, nor is it monotone on any subinterval of $[a, b]$.*

Using Lebesgue's monotone differentiation theorem (monotonicity implies differentiability almost everywhere), the second assertion follows from the first, but below we give a direct proof.

Before getting into the proofs of Baire's theorem and Theorem 32, we further discuss thickness, thinness, and genericity. The empty set is always thin and the full space M is always thick in itself. A single open dense subset is thick and a single closed nowhere dense subset is thin. $\mathbb{R} \setminus \mathbb{Z}$ is a thick subset of $\mathbb{R}$ and the Cantor set is a thin subset of $\mathbb{R}$. Likewise $\mathbb{R}$ is a

† "Residual" is an unfortunate choice of words. It connotes smallness, when it should connote just the opposite.

thin subset of $\mathbb{R}^2$. The generic point of $\mathbb{R}$ does not lie in the Cantor set. The generic point of $\mathbb{R}^2$ does not lie on the x-axis. Although $\mathbb{R} \setminus \mathbb{Z}$ is a thick subset of $\mathbb{R}$ it is not a thick subset of $\mathbb{R}^2$. The set $\mathbb{Q}$ is a thin subset of $\mathbb{R}$. It is the countable union of its points, each of which is a closed nowhere dense set. $\mathbb{Q}^c$ is a thick subset of $\mathbb{R}$. The generic real number is irrational. In the same vein:

(a) The generic square matrix has determinant $\neq 0$.
(b) The generic linear transformation $\mathbb{R}^m \to \mathbb{R}^m$ is an isomorphism.
(c) The generic linear transformation $\mathbb{R}^m \to \mathbb{R}^{m-k}$ is onto.
(d) The generic linear transformation $\mathbb{R}^m \to \mathbb{R}^{m+k}$ is one-to-one.
(e) The generic pair of lines in $\mathbb{R}^3$ are skew (nonparallel and disjoint).
(f) The generic plane in $\mathbb{R}^3$ meets the three coordinate axes in three distinct points.
(g) The generic n^{th} degree polynomial has n distinct roots.

In an incomplete metric space such as $\mathbb{Q}$, thickness and thinness have no bite: every subset of $\mathbb{Q}$, even the empty set, is thick in $\mathbb{Q}$.

Proof of Baire's Theorem If $M = \emptyset$, the proof is trivial, so we assume $M \neq \emptyset$. Let $G = \bigcap G_n$ be a thick subset of M; each G_n being open dense in M. Let $p_0 \in M$ and $\epsilon > 0$ be given. Choose a sequence of points $p_n \in M$ and radii $r_n > 0$ such that $r_n < 1/n$ and

$$\begin{aligned} M_{2r_1}(p_1) &\subset M_\epsilon(p_0) \\ M_{2r_2}(p_2) &\subset M_{r_1}(p_1) \cap G_1 \\ &\cdots \\ M_{2r_n}(p_{n+1}) &\subset M_{r_n}(p_n) \cap G_1 \cdots \cap G_n. \end{aligned}$$

See Figure 101. Then

$$M_\epsilon(p_0) \supset \overline{M}_{r_1}(p_1) \supset \overline{M}_{r_2}(p_2) \supset \ldots.$$

The diameters of these sets tend to 0 as $n \to \infty$. Thus (p_n) is a Cauchy sequence and it converges to some $p \in M$, by completeness. The point p belongs to each set $\overline{M}_{r_n}(p_n)$ and therefore it belongs to each G_n. Thus $p \in G \cap M_\epsilon(p_0)$ and G is dense in M.

To check that M is not thin, we take complements. Suppose that $M = \bigcup K_n$ and K_n is closed. If each K_n has empty interior, then each $G_n = K_n^c$ is open-dense, and

$$G = \bigcap G_n = (\bigcup K_n)^c = \emptyset,$$

a contradiction to density of G. □

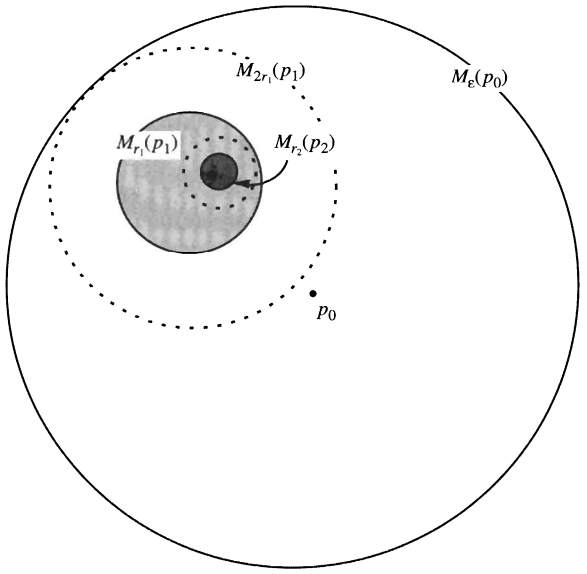


Figure 101 The closed neighborhoods $M_{r_n}(p_n)$ nest down to a point.

33 Corollary *No subset of a complete non-empty metric space is both thick and thin.*

Proof If S is both a thick and thin subset of M then $M \setminus S$ is also both thick and thin. The intersection of two thick subsets of M is thick, so $\emptyset = S \cap (M \setminus S)$ is a thick subset of M. By Baire's theorem, this empty set is dense in M, so M is empty. □

To prove Theorem 32 we use two lemmas.

34 Lemma *The set PL of piecewise linear functions is dense in C^0.*

Proof If $\phi : [a, b] \to \mathbb{R}$ is continuous and its graph consists of finitely many line segments in $\mathbb{R}^2$ then ϕ is **piecewise linear**. Let $f \in C^0$ and $\epsilon > 0$ be given. Since $[a, b]$ is compact, f is uniformly continuous, and there is $\delta > 0$ such that $|t - s| < \delta$ implies $|f(t) - f(s)| < \epsilon$. Choose $n > (b-a)/\delta$ and partition $[a, b]$ into n equal subintervals $I_i = [x_{i-1}, x_i]$, each of length $< \delta$. Let $\phi : [a, b] \to \mathbb{R}$ be the piecewise linear function whose graph consists of the segments joining the points $(x_{i-1}, f(x_{i-1}))$ and $(x_i, f(x_i))$ on the graph of f. See Figure 102.

The value of $\phi(t)$ for $t \in I_i$ lies between $f(x_{i-1})$ and $f(x_i)$. Both these numbers differ from $f(t)$ by less than ϵ. Hence, for all $t \in [a, b]$,

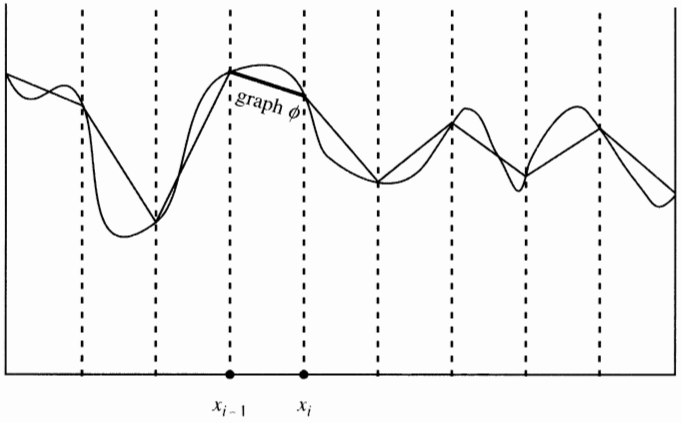


Figure 102 The piecewise linear function ϕ approximates the continuous function f.

$$|f(t) - \phi(t)| < \epsilon.$$

In other words, $d(f, \phi) < \epsilon$ and PL is dense in C^0. □

35 Lemma *If $\phi \in PL$ and $\epsilon > 0$ are given then there exists a sawtooth function σ such that $\|\sigma\| \leq \epsilon$, σ has period $\leq \epsilon$, and*

$$\min\{|\text{slope}\,\sigma|\} > \max\{|\text{slope}\,\phi|\} + \frac{1}{\epsilon}.$$

Proof Let $\theta = \max\{|\text{slope}\,\phi|\}$ and choose c large. The compressed sawtooth $\sigma(x) = \epsilon\sigma_0(cx)$ has $\|\sigma\| = \epsilon$, period $\pi = 1/c$, and slope $s = \pm\epsilon c$. When c is large, $\pi \leq \epsilon$, and $|s| > \theta + 1/\epsilon$. □

Proof of Theorem 32 For $n \in \mathbb{N}$ define

$$R_n = \{f \in C^0 : \forall x \in [a, b - \frac{1}{n}]\ \ \exists h > 0 \text{ such that } \left|\frac{\Delta f}{h}\right| > n\}$$

$$L_n = \{f \in C^0 : \forall x \in [a + \frac{1}{n}, b]\ \ \exists h < 0 \text{ such that } \left|\frac{\Delta f}{h}\right| > n\}$$

$$G_n = \{f \in C^0 : f \text{ restricted to any interval of length } \frac{1}{n} \text{ is non-monotone}\},$$

where $\Delta f = f(x + h) - f(x)$. We claim that each of these sets is open dense in C^0.

For denseness it is enough to prove that the closure of each contains PL, since by Lemma 34 the closure of PL is C^0. Let $\phi \in PL$ and $\epsilon > 0$ be given. According to Lemma 35, there is a sawtooth function σ such that $\|\sigma\| \leq \epsilon$, σ has period $< 1/n$, and

$$\min\{|\text{slope}\,\sigma|\} > \max\{|\text{slope}\,\phi|\} + n.$$

Consider the piecewise linear function $f = \phi + \sigma$. Its slopes are dominated by those of σ, and so they alternate in sign with period $< 1/2n$. At any $x \in [a, b + 1/n]$ there is a rightward slope either $> n$ or $< -n$. Thus $f \in R_n$. Similarly $f \in L_n$. Any interval I of length $1/n$ contains in its interior a maximum or minimum of σ, and so it contains a subinterval on which f strictly increases, and another on which f strictly decreases. Thus, $f \in G_n$. Since $d(f, \phi) = \epsilon$ is arbitrarily small this shows that R_n, L_n, and G_n are dense in C^0.

Next suppose that $f \in R_n$ is given. For each $x \in [a, b - 1/n]$ there is an $h = h(x) > 0$ such that

$$\left|\frac{f(x + h) - f(x)}{h}\right| > n.$$

Since f is continuous, there is a neighborhood T_x of x in $[a, b]$ and a constant $\nu = \nu > 0$ such that this same h yields

$$\left|\frac{f(t + h) - f(t)}{h}\right| > n + \nu$$

for all $t \in T_x$. Since $[a, b - 1/n]$ is compact, finitely many of these neighborhoods T_x cover it, say $T_{x_1}, \ldots, T_{x_m}$. Continuity of f implies that for all $t \in \overline{T}_{x_i}$,

$$\left|\frac{f(t + h_i) - f(t)}{h_i}\right| \geq n + \nu_i, \tag{25}$$

where $h_i = h(x_i)$ and $\nu_i = \nu(x_i)$. These m inequalities for points t in the m sets T_{x_i} remain nearly valid if f is replaced by a function g with $d(f, g)$ small enough; (25) becomes

$$\left|\frac{g(t + h_i) - g(t)}{h_i}\right| > n \tag{26}$$

which means that $g \in R_n$ and R_n is open in C^0. Similarly L_n is open in C^0.

Checking that G_n is open is easier. If (f_k) is a sequence of functions in G_n^c and $f_k \rightrightarrows f$ then we must show that $f \in G_n^c$. Each f_k is monotone on some interval I_k of length $1/n$. There is a subsequence of these intervals that converges to a limit interval I. Its length is $1/n$ and by uniform convergence, f is monotone on I. Hence G_n^c is closed and G_n is open. Each set R_n, L_n, G_n is open dense in C^0.

Finally, if f belongs to the thick set

$$\bigcap R_n \cap L_n \cap G_n$$

then for each $x \in [a, b]$ there is a sequence $h_n \neq 0$ such that

$$\left| \frac{f(x + h_n) - f(x)}{h_n} \right| > n.$$

The numerator of this fraction is at most $2 \| f \|$, so $h_n \to 0$ as $n \to \infty$. Thus f is not differentiable at x. Also, f is non-monotone on every interval of length $1/n$. Since each interval J contains an interval of length $1/n$ when n is large enough, f is non-monotone on J. □

Further generic properties of continuous functions have been studied, and you might read about them in the books *A Primer of Real Functions* by Ralph Boas, *Differentiation of Real Functions* by Andrew Bruckner, or *A Second Course in Real Functions* by van Rooij and Schikhof.

8* Spaces of Unbounded Functions

How important is it that the functions we deal with are bounded, or have domain $[a, b]$ and target $\mathbb{R}$? To some extent we can replace $[a, b]$ with a metric space X and $\mathbb{R}$ with a complete metric space Y. Let $\mathcal{F}$ denote the set of all functions $f : X \to Y$. Recall from Exercise 2.84 that the metric d_Y on Y gives rise to a bounded metric

$$\rho(y, y') = \frac{d_Y(y, y')}{1 + d_Y(y, y')},$$

where $y, y' \in Y$. Note that $\rho < 1$. Convergence and Cauchyness with respect to ρ and d_Y are equivalent. Thus completeness of Y with respect to d_Y implies completeness with respect to ρ. In the same way we give $\mathcal{F}$ the metric

$$d(f, g) = \sup_{x \in X} \frac{d_Y(f(x), g(x))}{1 + d_Y(f(x), g(x))}.$$

A function $f \in \mathcal{F}$ is **bounded** with respect to d_Y if and only if for any constant function c, $\sup_x d_Y(f(x), c) < \infty$; i.e., $d(f, c) < 1$. Unbounded functions have $d(f, c) = 1$.

36 Theorem *In the space $\mathcal{F}$ equipped with the metric d,*
(a) Uniform convergence of (f_n) is equivalent to d-convergence.
(b) Completeness of Y implies completeness of $\mathcal{F}$.
(c) The set $\mathcal{F}_b$ of bounded functions is closed in $\mathcal{F}$.
(d) The set $C^0(X, Y)$ of continuous functions is closed in $\mathcal{F}$.

Proof (a) $f = \operatorname{unif\,lim}_{n \to \infty}$ means that $d_Y(f_n(x), f(x)) \rightrightarrows 0$, which means that $d(f_n, f) \to 0$.

(b) If (f_n) is Cauchy in $\mathcal{F}$ and Y is complete then, just as in Section 1, $f(x) = \lim_{n \to \infty} f_n(x)$ exists for each $x \in X$. Cauchyness with respect to the metric d implies uniform convergence and thus $d(f_n, f) \to 0$.

(c) If $f_n \in \mathcal{F}_b$ and $d(f_n, f) \to 0$ then $\sup_x d_Y(f_n(x), f(x)) \to 0$. Since f_n is bounded, so is f.

(d) The proof that C^0 is closed in $\mathcal{F}$ is the same as in Section 1. □

The Arzela-Ascoli theorem is trickier. A family $\mathcal{E} \subset \mathcal{F}$ is **uniformly equicontinuous** if for each $\epsilon > 0$ there is a $\delta > 0$ such that $f \in \mathcal{E}$ and $d_X(x, t) < \delta$ imply $d_Y(f(x), f(t)) < \epsilon$. If the δ depends on x but not on $f \in \mathcal{E}$ then $\mathcal{E}$ is **pointwise equicontinuous**.

37 Theorem *Pointwise equicontinuity implies uniform equicontinuity if X is compact.*

Proof Suppose not. Then there exists $\epsilon > 0$ such that for each $\delta = 1/n$ we have points $x_n, t_n \in X$ and functions $f_n \in \mathcal{E}$ with $d_X(x_n, t_n) < 1/n$ and $d_Y(f_n(x_n), f_n(t_n)) \geq \epsilon$. By compactness of X we may assume that $x_n \to x_0$. Then $t_n \to x_0$, which leads to a contradiction of pointwise equicontinuity at x_0. □

38 Theorem *If the sequence of functions $f_n : X \to Y$ is uniformly equicontinuous, X is compact, and for each $x \in X$, $(f_n(x))$ lies in a compact subset of Y, then (f_n) has a uniformly convergent subsequence.*

Proof Being compact, X has a countable dense subset D. Then the proof of the Arzela Ascoli Theorem in Section 3 becomes a proof of Theorem 38. □

The space X is **σ-compact** if it is a countable union of compact sets, $X = \bigcup X_i$. For example $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{R}^m$ are σ-compact, while any uncountable set equipped with the discrete metric is not σ-compact.

39 Theorem *If X is σ-compact and if (f_n) is a sequence of pointwise equicontinuous functions such that for each $x \in X$, $(f_n(x))$ lies in a compact subset of Y, then (f_n) has a subsequence that converges uniformly to a limit on each compact subset of X.*

Proof Express X as $\bigcup X_i$ with X_i compact. By Theorem 37 $(f_n|_{X_i})$ is uniformly equicontinuous and by Theorem 38 there is a subsequence $f_{1,n}$ that converges uniformly on X_1, and it has a sub-subsequence $f_{2,n}$ that converges uniformly on X_2, and so on. A diagonal subsequence (g_m) converges uniformly on each X_i. Thus (g_m) converges pointwise. If $A \subset X$ is compact, then $(g_m|_A)$ is uniformly equicontinuous and pointwise convergent. By the proof of the Arzela Ascoli propagation theorem, $(g_m|_A)$ converges uniformly. □

40 Corollary *If (f_n) is a sequence of pointwise equicontinuous functions $\mathbb{R} \to \mathbb{R}$, and for some $x_0 \in \mathbb{R}$, $(f_n(x_0))$ is bounded then (f_n) has a subsequence that converges uniformly on every compact subset of $\mathbb{R}$.*

Proof Let $[a, b]$ be any interval containing x_0. By Theorem 37, the restrictions of f_n to $[a, b]$ are uniformly equicontinuous, and there is a $\delta > 0$ such that if $t, s \in [a, b]$ then $|t - s| < \delta$ implies that $|f_n(t) - f_n(s)| < 1$. Each point $x \in [a, b]$ can be reached in $\leq N$ steps of length $< \delta$, starting at x_0, if $N > (b - a)/\delta$. Thus $|f_n(x)| \leq |f_n(x_0)| + N$, and $(f_n(x))$ is bounded for each $x \in \mathbb{R}$. A bounded subset of $\mathbb{R}$ has compact closure and Theorem 39 gives the corollary. □

Exercises

In these exercises, $C^0 = C^0([a, b], \mathbb{R})$ is the space of continuous real valued functions defined on the closed interval $[a, b]$. It is equipped with the sup norm, $\|f\| = \sup\{|f(x)| : x \in [a, b]\}$.

1. Let M, N be metric spaces.
   (a) Formulate the concepts of pointwise convergence and uniform convergence for sequences of functions $f_n : M \to N$.
   (b) For which metric spaces are the concepts equivalent?
2. Suppose that $f_n \rightrightarrows f$ where f and f_n are functions from the metric space M to the metric space N. (Assume nothing about the metric spaces such as compactness, completeness, etc.) If each f_n is continuous prove that f is continuous. [Hint: Review the proof of Theorem 1.]
3. Let $f_n : [a, b] \to \mathbb{R}$ be a sequence of piecewise continuous functions, each of which is continuous at the point $x_0 \in [a, b]$. Assume that $f_n \rightrightarrows f$ as $n \to \infty$.
   (a) Prove that f is continuous at x_0. [Hint: Review the proof of Theorem 1.]
   (b) Prove or disprove that f is piecewise continuous.
4. (a) If $f_n : \mathbb{R} \to \mathbb{R}$ is uniformly continuous for each $n \in \mathbb{N}$ and if $f_n \rightrightarrows f$ as $n \to \infty$, prove or disprove that f is uniformly continuous.
   (b) What happens for functions from one metric space to another instead of $\mathbb{R}$ to $\mathbb{R}$?
5. Suppose that $f_n : [a, b] \to \mathbb{R}$ and $f_n \rightrightarrows f$ as $n \to \infty$. Which of the following discontinuity properties (see Exercise 37 in Chapter 3) of the functions f_n carries over to the limit function? (Prove or give a counter-example.)
   (a) No discontinuities.
   (b) At most ten discontinuities.
   (c) At least ten discontinuities.
   (d) Finitely many discontinuities.
   (e) Countably many discontinuities, all of jump type.
   (f) No jump discontinuities.
   (g) No oscillating discontinuities.

**6. (a) Prove that C^0 and $\mathbb{R}$ have equal cardinality. [Clearly there are at least as many functions as there are real numbers, for C^0 includes the constant functions. The issue is to show that there are no more continuous functions than there are real numbers.]

(b) Is the same true if we replace $[a, b]$ with $\mathbb{R}$ or a separable metric space?

(c) In the same vein, prove that the collection $\mathcal{T}$ of open subsets of $\mathbb{R}$ and $\mathbb{R}$ itself have equal cardinality.

7. Consider a sequence of functions f_n in C^0. The graph G_n of f_n is a compact subset of $\mathbb{R}^2$.
   (a) Prove that (f_n) converges uniformly as $n \to \infty$ if and only if the sequence (G_n) in $\mathcal{K}(\mathbb{R}^2)$ converges to the graph of a function $f \in C^0$. (The space $\mathcal{K}$ was discussed in Exercise 2.124.)
   (b) Formulate equicontinuity in terms of graphs.
8. Is the sequence of functions $f_n : \mathbb{R} \to \mathbb{R}$ defined by

$$f_n(x) = \cos(n + x) + \log(1 + \frac{1}{\sqrt{n+2}} \sin^2(n^n x))$$

equicontinuous? Prove or disprove.

9. If $f : \mathbb{R} \to \mathbb{R}$ is continuous and the sequence $f_n(x) = f(nx)$ is equicontinuous, what can be said about f?
10. Give an example to show that a sequence of functions may be uniformly continuous, pointwise equicontinuous, but not uniformly equicontinuous, when their domain M is noncompact.
11. If every sequence of pointwise equicontinuous functions $M \to \mathbb{R}$ is uniformly equicontinuous, does this imply that M is compact?
12. Prove that if $\mathcal{E} \subset C^0(M, N)$ is equicontinuous then so is its closure.
13. Suppose that (f_n) is a sequence of functions $\mathbb{R} \to \mathbb{R}$ and for each compact subset $K \subset \mathbb{R}$, the restricted sequence $(f_n|_K)$ is pointwise bounded and pointwise equicontinuous.
   (a) Does it follow that there is a subsequence of (f_n) that converges pointwise to a continuous limit function $\mathbb{R} \to \mathbb{R}$?
   (b) What about uniform convergence?
14. Recall from Exercise 78 in Chapter 2 that a metric space M is chain connected if for each $\epsilon > 0$ and each $p, q \in M$ there is a chain $p = p_0, \ldots, p_n = q$ in M such that

$$d(p_{k-1}, p_k) < \epsilon \quad \text{for} \quad 1 \le k \le n.$$

A family $\mathcal{F}$ of functions $f : M \to \mathbb{R}$ is bounded at $p \in M$ if the set $\{f(p) : f \in \mathcal{F}\}$ is a bounded in $\mathbb{R}$.
Show that M is chain connected if and only if pointwise boundedness of an equicontinuous family at one point of M implies pointwise boundedness at every point of M.

15. A continuous, strictly increasing function $\mu : (0, \infty) \to (0, \infty)$ is a **modulus of continuity** if $\mu(s) \to 0$ as $s \to 0$. A function $f : [a, b] \to \mathbb{R}$ has modulus of continuity μ if $|f(s) - f(t)| \le \mu(|s - t|)$.
   (a) Prove that a function is uniformly continuous if and only if it has a modulus of continuity.
   (b) Prove that a family of functions is equicontinuous if and only if its members have a common modulus of continuity.
16. Consider the modulus of continuity $\mu(s) = Ls$ where L is a positive constant.
   (a) What is the relation between C^μ and the set of Lipschitz functions with Lipschitz constant $\le L$?
   (b) Replace $[a, b]$ with $\mathbb{R}$ and answer the same question.
   (c) Replace $[a, b]$ with $\mathbb{N}$ and answer the same question.
   (d) Formulate and prove a generalization of (a).
17. Consider a modulus of continuity $\mu(s) = Hs^\alpha$ where $0 < \alpha \le 1$ and $0 < H < \infty$. A function with this modulus of continuity is said to be **α-Hölder**, with α-Hölder constant H. See also Exercise 2 in Chapter 3.
   (a) Prove that the set $C^\alpha(H)$ of all continuous functions defined on $[a, b]$ which are α-Hölder and have α-Hölder constant $\le H$ is equicontinuous.
   (b) Replace $[a, b]$ with (a, b). Is the same thing true?
   (c) Replace $[a, b]$ with $\mathbb{R}$. Is it true?
   (d) What about $\mathbb{Q}$?
   (e) What about $\mathbb{N}$?
18. Suppose that (f_n) is an equicontinuous sequence in C^0 and $p \in [a, b]$ is given.
   (a) If $f_n(p)$ is a bounded sequence of real numbers, prove that (f_n) is uniformly bounded.
   (b) Reformulate the Arzela-Ascoli Theorem with a weaker boundedness hypothesis.
   (c) Can $[a, b]$ be replaced with (a, b)?, $\mathbb{Q}$?, $\mathbb{R}$?, $\mathbb{N}$?
   (d) What is the correct generalization?
19. If M is compact and A is dense in M, prove that for any $\delta > 0$ there is a finite subset $\{a_1, \ldots, a_k\} \subset A$ which is **δ-dense** in M in the sense that each $x \in M$ lies within distance δ of at least one of the points a_j, $j = 1, \ldots, k$.

20. Give an example of a sequence of smooth equicontinuous functions $f_n : [a, b] \to \mathbb{R}$ whose derivatives are unbounded.
21. Suppose that (f_n) is an equicontinuous sequence of functions $f_n : \mathbb{R} \to \mathbb{R}$, such that $(f_n(0))$ is a bounded sequence in $\mathbb{R}$.
    (a) Does there exist a subsequence that converges uniformly? Prove or give a counter-example.
    (b) What if $\mathbb{R}^m$ replaces $\mathbb{R}$?
    (c) What about $\mathbb{N}$?
22. Suppose that $\mathcal{E} \subset C^0$ is equicontinuous and bounded.
    (a) Prove that $\sup\{f(x) : f \in \mathcal{E}\}$ is a continuous function of x.
    (b) Show that (a) fails without equicontinuity.
    (c) Show that this continuous sup-property does not imply equicontinuity.
    (d) Assume that the continuous sup-property is true for each subset $\mathcal{F} \subset \mathcal{E}$. Is $\mathcal{E}$ equicontinuous? Give a proof or counter-example.
23. Let M be a compact metric space, and let (i_n) be a sequence of isometries $i_n : M \to M$.
    (a) Prove that there exists a subsequence i_{n_k} that converges to an isometry i as $k \to \infty$.
    (b) Does the inverse isometry $i_{n_k}^{-1}$ converge to i^{-1}. (Proof or counter-example.)
    (c) Infer that the group of orthogonal 3×3 matrices is compact. [Hint: Each orthogonal 3×3 matrix defines an isometry of the unit 2-sphere to itself.]
    (d) How about the group of $m \times m$ orthogonal matrices?
24. Suppose that $f : M \to M$ is a contraction, but M is not necessarily complete.
    (a) Prove that f is uniformly continuous.
    (b) Why does (a) imply that f extends uniquely to a continuous map $\widehat{f} : \widehat{M} \to \widehat{M}$, where $\widehat{M}$ is the completion of M?
    (c) Is $\widehat{f}$ a contraction?
25. Give an example of a contraction of an incomplete metric space that has no fixed point.
26. Suppose that $f : M \to M$ and for all $x, y \in M$, if $x \neq y$ then $d(f(x), f(y)) < d(x, y)$. Such an f is a **weak contraction**.
    (a) Is a weak contraction a contraction? (Proof or counter-example.)
    (b) If M is compact is a weak contraction a contraction? (Proof or counter-example.)

    (c) If M is compact, prove that a weak contraction has a unique fixed point.
27. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is differentiable and its derivative satisfies $|f'(x)| < 1$ for all $x \in \mathbb{R}$.
    (a) Is f a contraction?
    (b) A weak one?
    (c) Does it have a fixed point?
28. Give an example to show that the fixed point in Brouwer's Theorem need not be unique.
29. On page 222 it is shown that if $b_n \int_{-1}^{1} (1 - t^2)^n \, dt = 1$ then for some constant c, and for all $n \in \mathbb{N}$, $b_n \leq c\sqrt{n}$. What is the best (i.e., smallest) value of c that you can prove works? (A calculator might be useful here.)
30. Let M be a compact metric space, and let C^{Lip} be the set of continuous functions $f : M \to \mathbb{R}$ that obey a Lipschitz condition: for some L and all $p, q, \in M$,

$$|f(p) - f(q)| \leq Ld(p, q).$$

    *(a) Prove that C^{Lip} is dense in $C^0(M, \mathbb{R})$. [Hint: Stone-Weierstrass.]
    ***(b) If $M = [a, b]$ and $\mathbb{R}$ is replaced by some other complete metric space, is the result true or false?
    ***(c) If M is a general compact metric space and Y is a complete metric space, is $C^{\text{Lip}}(M, Y)$ dense in $C^0(M, Y)$? (Would M equal to the Cantor set make a good test case?)
31. Consider the ODE $x' = x$ on $\mathbb{R}$. Show that its solution with initial condition x_0 is $t \mapsto e^t x_0$. Interpret $e^{t+s} = e^t e^s$ in terms of the flow property.
32. Consider the ODE $y' = 2\sqrt{|y|}$ where $y \in \mathbb{R}$.
    (a) Show that there are many solutions to this ODE, all with the same initial condition $y(0) = 0$. Not only does $y(t) = 0$ solve the ODE, but also $y(t) = t^2$ does for $t \geq 0$.
    (b) Find and graph other solutions such as $y(t) = 0$ for $t \leq c$ and $y(t) = (t - c)^2$ for $t \geq c > 0$.
    (c) Does the existence of these non-unique solutions to the ODE contradict Picard's Theorem? Explain.
    *(d) Find all solutions with initial condition $y(0) = 0$.
33. Consider the ODE $x' = x^2$ on $\mathbb{R}$. Find the solution of the ODE with initial condition x_0. Are the solutions to this ODE defined for all time or do they escape to infinity in finite time?

34. Suppose that the ODE $x' = f(x)$ on $\mathbb{R}$ is bounded, $|f(x)| \leq M$ for all x.
    (a) Prove that no solution of the ODE escapes to infinity in finite time.
    (b) Prove the same thing if f satisfies a Lipschitz condition, or, more generally, if there are constants C, K such that $|f(x)| \leq C|x| + K$ for all x.
    (c) Repeat (a) and (b) with $\mathbb{R}^m$ in place of $\mathbb{R}$.
    (d) Prove that if $f : \mathbb{R}^m \to \mathbb{R}^m$ is uniformly continuous then the condition stated in (b) is true. Deduce that solutions of uniformly continuous ODE's defined on $\mathbb{R}^m$ do not escape to infinity in finite time.

35. (a) Prove **Borel's Lemma: given any sequence whatsoever of real numbers (a_r) there is a smooth function $f : \mathbb{R} \to \mathbb{R}$ such that $f^{(r)}(0) = a_r$. [Hint: Try $f = \sum \beta_k(x) a_k x^k / k!$ where β_k is a well chosen bump function.]
    (b) Infer that there are many Taylor series with radius of convergence $R = 0$.
    (c) Construct a smooth function whose Taylor series at every x has radius of convergence $R = 0$. [Hint: Try $\sum \beta_k(x) e(x + q_k)$ where $\{q_1, q_2, \dots\} = \mathbb{Q}$.]

*36. Suppose that $T \subset (a, b)$ clusters at some point of (a, b) and that $f, g : (a, b) \to \mathbb{R}$ are analytic. Assume that for all $t \in T$, $f(t) = g(t)$.
    (a) Prove that $f = g$ everywhere in (a, b).
    (b) What if f and g are only C^∞?
    (c) What if T is an infinite set but its only cluster points are a and b?
    (d) Find a necessary and sufficient condition for a subset $Z \subset (a, b)$ to be the **zero locus of an analytic function f defined on (a, b), $Z = \{x \in (a, b) : f(x) = 0\}$. [Hint: Think Taylor. The result in (a) is known as the **Identity Theorem**. It states that if an equality between analytic functions is known to hold for points of T then it is an **identity**, an equality that holds everywhere.]

37. Let M be any metric space with metric d. Fix a point $p \in M$ and for each $q \in M$ define the function $f_q(x) = d(q, x) - d(p, x)$.
    (a) Prove that f_q is a bounded, continuous function of $x \in M$, and that the map $q \mapsto f_q$ sends M isometrically onto a subset M_0 of $C^0(M, \mathbb{R})$.

    (b) Since $C^0(M, \mathbb{R})$ is complete, infer that an isometric copy of M is dense in a complete metric space, the closure of M_0, and hence that we have a second proof of the Completion Theorem 2.76.

38. As explained in Section 8, a metric space M is **σ-compact** if it is the countable union of compact subsets, $M = \bigcup M_i$.
    (a) Why is it equivalent to require that M is the monotone union of compact subsets,

$$M = \bigcup\!\!\uparrow M_i \ ,$$

    i.e., $M_1 \subset M_2 \subset \dots$?
    (b) Prove that a σ-compact metric space is separable.
    (c) Prove that $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{R}^m$ are σ-compact.
    *(d) Prove that C^0 is not σ-compact. [Hint: Think Baire.]
    *(e) If $M = \bigcup\!\!\uparrow \operatorname{int}(M_i)$ and each M_i is compact, M is **σ^*-compact**. Prove that M is σ^*-compact if and only if it is separable and locally compact. Infer that $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{R}^m$ are σ^*-compact, but $\mathbb{Q}$ is not.
    (f) Assume that M is σ^*-compact, $M = \bigcup\!\!\uparrow \operatorname{int}(M_i)$ with each M_i compact. Prove that this monotone union **engulfs** all compacts in M, in the sense that if $A \subset M$ is compact, then for some i, $A \subset M_i$.
    (g) If $M = \bigcup\!\!\uparrow M_i$ and each M_i is compact show by example that this engulfing property may fail, even when M itself is compact.
    **(h) Prove or disprove that a complete σ-compact metric space is σ^*-compact.

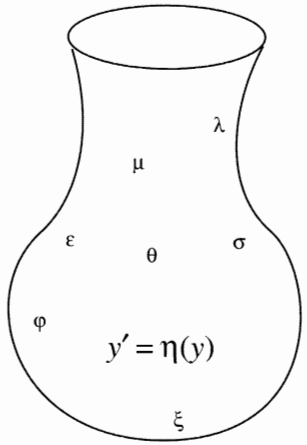
39. (a) Give an example of a function $f : [0, 1] \times [0, 1] \to \mathbb{R}$ such that for each fixed x, $y \mapsto f(x, y)$ is a continuous function of y, and for each fixed y, $x \mapsto f(x, y)$ is a continuous function of x, but f is not continuous.
    (b) Suppose in addition that the set of functions

$$\mathcal{E} = \{x \mapsto f(x, y) : y \in [0, 1]\}$$

    is equicontinuous. Prove that f is continuous.

40. Prove that $\mathbb{R}$ can not be expressed as the countable union of Cantor sets.

41. What is the joke in the following picture?



More Pre-lim Problems

1. Let f and $f_n, n \in \mathbb{N}$, be functions from $\mathbb{R}$ to $\mathbb{R}$. Assume that $f_n(x_n) \to f(x)$ as $n \to \infty$ and $x_n \to x$. Show that f is continuous. [Hint: Equicontinuity is irrelevant since the functions f_n are not assumed to be continuous.]
2. Suppose that $f_n \in C^0$ and for each $x \in [a, b]$,

$$f_1(x) \geq f_2(x) \geq \ldots,$$

and $\lim_{n\to\infty} f_n(x) = 0$. Is the sequence equicontinuous? Give a proof or counter-example. [Hint: Does $f_n(x)$ converge uniformly to 0, or does it not?]
3. Let E be the set of all functions $u : [0, 1] \to \mathbb{R}$ such that $u(0) = 0$ and u satisfies a Lipschitz condition with Lipschitz constant 1. Define $\phi : E \to \mathbb{R}$ according to the formula

$$\phi(u) = \int_0^1 (u(x)^2 - u(x))\, dx.$$

Prove that there exists a function $u \in E$ at which $\phi(u)$ attains an absolute maximum.
4. Let (g_n) be a sequence of twice differentiable functions defined on $[0, 1]$, and assume that for all n, $g_n(0) = g_n'(0)$. Suppose also that for all $n \in \mathbb{N}$ and all $x \in [0, 1]$, $\left|g_n'(x)\right| \leq 1$. Prove that there is a subsequence of (g_n) converging uniformly on $[0, 1]$.

5. Let (P_n) be a sequence of real polynomials of degree ≤ 10. Suppose that $P_n(x)$ converges pointwise to 0 as $n \to \infty$ and $x \in [0, 1]$. Prove that $P_n(x)$ converges uniformly to 0.
6. Let (a_n) be a sequence of nonzero real numbers. Prove that the sequence of functions

$$f_n(x) = \frac{1}{a_n} \sin(a_n x) + \cos(x + a_n)$$

has a subsequence converging to a continuous function.
7. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is differentiable, $f(0) = 0$, and $f'(x) > f(x)$ for all $x \in \mathbb{R}$. Prove that $f(x) > 0$ for all $x > 0$.
8. Suppose that $f : [a, b] \to \mathbb{R}$ and the limits of $f(x)$ from the left and the right exist at all points of $[a, b]$. Prove that f is Riemann integrable.
9. Let $h : [0, 1) \to \mathbb{R}$ be a uniformly continuous function where $[0, 1)$ is the half open interval. Prove that there is a unique continuous map $g : [0, 1] \to \mathbb{R}$ such that $g(x) = h(x)$ for all $x \in [0, 1)$.
10. Assume that $f : \mathbb{R} \to \mathbb{R}$ is uniformly continuous. Prove that there are constants A, B such that $|f(x)| \leq A + B\,|x|$ for all $x \in \mathbb{R}$.
11. Suppose that $f(x)$ is defined on $[-1, 1]$ and that its third derivative exists and is continuous. (That is, f is of class C^3.) Prove that the series

$$\sum_{n=0}^{\infty} \Big(n(f(1/n) - f(-1/n)) - 2f'(0)\Big)$$

converges.
12. Let $A \subset \mathbb{R}^m$ be compact, $x \in A$. Let (x_n) be a sequence in A such that every convergent subsequence of (x_n) converges to x.
    (a) Prove that the sequence (x_n) converges.
    (b) Give an example to show if A is not compact, the result in (a) is not necessarily true.
13. Let $f : [0, 1] \to \mathbb{R}$ be continuously differentiable, with $f(0) = 0$. Prove that

$$\|f\|^2 \leq \int_0^1 (f'(x))^2\, dx$$

where $\|f\| = \sup\{|f(t)| : 0 \leq t \leq 1\}$.
14. Let $f_n : \mathbb{R} \to \mathbb{R}$ be differentiable functions, $n = 1, 2, \ldots$ with $f_n(0) = 0$ and $\left|f_n'(x)\right| \leq 2$ for all n, x. Suppose that

$$\lim_{n\to\infty} f_n(x) = g(x)$$

for all x. Prove that g is continuous.

15. Let X be a non-empty connected set of real numbers. If every element of X is rational, prove that X has only one element.
16. Let $k \geq 0$ be an integer and define a sequence of maps $f_n : \mathbb{R} \to \mathbb{R}$ as

$$f_n(x) = \frac{x^k}{x^2 + n}$$

$n = 1, 2, \ldots$. For which values of k does the sequence converge uniformly on $\mathbb{R}$? On every bounded subset of $\mathbb{R}$?
17. Let $f : [0, 1] \to \mathbb{R}$ be Riemann integrable over $[b, 1]$ for every b such that $0 < b \leq 1$.
    (a) If f is bounded, prove that f is Riemann integrable over $[0, 1]$.
    (b) What if f is not bounded?
18. (a) Let S and T be connected subsets of the plane $\mathbb{R}^2$ having a point in common. Prove that $S \cup T$ is connected.
    (b) Let $\{S_\alpha\}$ be a family of connected subsets of $\mathbb{R}^2$ all containing the origin. Prove that $\bigcup S_\alpha$ is connected.
19. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous. Suppose that $\mathbb{R}$ contains a countably infinite set S such that

$$\int_p^q f(x)\,dx = 0$$

if p and q are not in S. Prove that f is identically zero.
20. Let $f : \mathbb{R} \to \mathbb{R}$ satisfy $f(x) \leq f(y)$ for $x \leq y$. Prove that the set where f is not continuous is finite or countably infinite.
21. Let (g_n) be a sequence of Riemann integrable functions from $[0, 1]$ into $\mathbb{R}$ such that $|g_n(x)| \leq 1$ for all n, x. Define

$$G_n(x) = \int_0^x g_n(t)\,dt.$$

Prove that a subsequence of (G_n) converges uniformly.
22. Prove that every compact metric space has a countable dense subset.
23. Show that for any continuous function $f : [0, 1] \to \mathbb{R}$ and for any $\epsilon > 0$ there is a function of the form

$$g(x) = \sum_{k=0}^{n} C_k x^k$$

for some $n \in \mathbb{N}$, and $|g(x) - f(x)| < \epsilon$ for all x in $[0, 1]$.
24. Give an example of a function $f : \mathbb{R} \to \mathbb{R}$ having all three of the following properties:

    (a) $f(x) = 0$ for all $x < 0$ and $x > 2$.
    (b) $f'(1) = 1$,
    (c) f has derivatives of all orders.
25. (a) Give an example of a differentiable function $f : \mathbb{R} \to \mathbb{R}$ whose derivative is not continuous.
    (b) Let f be as in (a). If $f'(0) < 2 < f'(1)$ prove that $f'(x) = 2$ for some $x \in [0, 1]$.
26. Let $U \subset \mathbb{R}^m$ be an open set. Suppose that the map $h : U \to \mathbb{R}^m$ is a homeomorphism from U onto $\mathbb{R}^m$ which is uniformly continuous. Prove that $U = \mathbb{R}^m$.
27. Let (f_n) be a sequence of continuous maps $[0, 1] \to \mathbb{R}$ such that

$$\int_0^1 (f_n(y))^2\,dy \leq 5$$

for all n. Define $g_n : [0, 1] \to \mathbb{R}$ by

$$g_n(x) = \int_0^1 \sqrt{x + y}\, f_n(y)\,dy$$

    (a) Find a constant $K \geq 0$ such that $|g_n(x)| \leq K$ for all n.
    (b) Prove that a subsequence of the sequence (g_n) converges uniformly.
28. Consider the following properties of a map $f : \mathbb{R}^m \to \mathbb{R}$.
    (a) f is continuous.
    (b) The graph of f is connected in $\mathbb{R}^m \times \mathbb{R}$.

    Prove or disprove the implications (a) $\Rightarrow$ (b), (b) $\Rightarrow$ (a).
29. Let (P_n) be a sequence of real polynomials of degree ≤ 10. Suppose that

$$\lim_{n\to\infty} P_n(x) = 0$$

for all $x \in [0, 1]$. Prove that $P_n(x) \rightrightarrows 0$, $0 \leq x \leq 1$. What can you say about $P_n(x)$ for $4 \leq x \leq 5$?
30. Give an example of a subset of $\mathbb{R}$ having uncountably many connected components. Can such a subset be open? Closed? Does your answer change if $\mathbb{R}^2$ replaces $\mathbb{R}$?
31. For each $(a, b, c) \in \mathbb{R}^3$ consider the series

$$\sum_{n=3}^{\infty} \frac{a^n}{n^b (\log n)^c}.$$

Determine the values of a, b, and c for which the series converges absolutely, converges conditionally, diverges.

32. Let X be a compact metric space and $f : X \to X$ an isometry. (That is, $d(f(x), f(y)) = d(x, y)$ for all $x, y \in X$.) Prove that $f(X) = X$.
33. Prove or disprove: $\mathbb{Q}$ is the countable intersection of open subsets of $\mathbb{R}$.
34. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and

$$\int_{-\infty}^{\infty} |f(x)|\, dx < \infty.$$

Show that there is a sequence (x_n) in $\mathbb{R}$ such that $x_n \to \infty$, $x_n f(x_n) \to 0$, and $x_n f(-x_n) \to 0$ as $n \to \infty$.

35. Let $f : [0, 1] \to \mathbb{R}$ be a continuous function. Evaluate the following limits (with proof)

(a) $\lim_{n\to\infty} \int_0^1 x^n f(x)\, dx$ (b) $\lim_{n\to\infty} n \int_0^1 x^n f(x)\, dx.$

36. Let K be an uncountable subset of $\mathbb{R}^m$. Prove that there is a sequence of distinct points in K which converges to some point of K.
37. Let (g_n) be a sequence of twice differentiable functions on $[0, 1]$ such that for all n, $g_n(0) = 0$ and $g_n'(0) = 0$. Suppose that $|g_n''(x)| \leq 1$ for all n, x. Prove that there is a subsequence of (g_n) which converges uniformly on $[0, 1]$.
38. Prove or give a counter-example: Every connected locally pathwise connected set in $\mathbb{R}^m$ is pathwise connected.
39. Let (f_n) be a sequence of continuous functions $[0, 1] \to \mathbb{R}$ such that $f_n(x) \to 0$ for each $x \in [0, 1]$. Suppose that

$$\left| \int_0^1 f_n(x)\, dx \right| \leq K$$

for all n where K is a constant. Does $\int_0^1 f_n(x)\, dx$ converge to 0 as $n \to \infty$? Prove or give a counter-example.

40. Let E be a closed, bounded, and non-empty subset of $\mathbb{R}^m$ and let $f : E \to E$ be a function satisfying $|f(x) - f(y)| < |x - y|$ for all $x, y \in E$, $x \neq y$. Prove that there is one and only one point $x_0 \in E$ such that $f(x_0) = x_0$.
41. Let $f : [0, 2\pi] \to \mathbb{R}$ be a continuous function such that

$$\int_0^{2\pi} f(x) \sin(nx)\, dx = 0$$

for all integers $n \geq 1$. Prove that f is identically zero.

42. Let E be the set of all real valued functions $u : [0, 1] \to \mathbb{R}$ satisfying $u(0) = 0$ and $|u(x) - u(y)| \leq |x - y|$ for all $x, y \in [0, 1]$. Prove that the function

$$\phi(u) = \int_0^1 ((u(x)^2 - u(x))\, dx$$

achieves its maximum value at some element of E.

43. Let $f_1, f_2, \ldots$ be continuous real valued functions on $[0, 1]$ such that for each $x \in [0, 1]$, $f_1(x) \geq f_2(x) \geq \ldots$. Assume that for each x, $f_n(x)$ converges to 0 as $n \to \infty$. Does f_n converge uniformly to 0? Give a proof or counter-example.
44. Let $f : [0, \infty) \to [0, \infty)$ be a monotonically decreasing function with

$$\int_0^{\infty} f(x)\, dx < \infty.$$

Prove that $\lim_{x\to\infty} xf(x) = 0$.

45. Suppose that $F : \mathbb{R}^m \to \mathbb{R}^m$ is continuous and satisfies

$$|F(x) - F(y)| \geq \lambda |x - y|$$

for all $x, y \in \mathbb{R}^m$ and some constant $\lambda > 0$. Prove that F is one-to-one, onto, and it has a continuous inverse.

46. Show that $[0, 1]$ cannot be written as a countably infinite union of disjoint closed sub-intervals.
47. Prove that a continuous function $f : \mathbb{R} \to \mathbb{R}$ which sends open sets to open sets must be monotonic.
48. Let $f : [0, \infty) \to \mathbb{R}$ be uniformly continuous and assume that

$$\lim_{b\to\infty} \int_0^b f(x)\, dx$$

exists (as a finite limit). Prove that $\lim_{x\to\infty} f(x) = 0$.

49. Prove or supply a counter-example: If f and g are continuously differentiable functions defined on the interval $0 < x < 1$ which satisfy the conditions

$$\lim_{x\to 0} f(x) = 0 = \lim_{x\to 0} g(x) \qquad \text{and} \qquad \lim_{x\to 0} \frac{f(x)}{g(x)} = c$$

and if g and g' never vanish, then $\lim_{x\to 0} \frac{f'(x)}{g'(x)} = c$. (This is a converse of L'Hospital's rule.)

50. Prove or provide a counter-example: If the function f from $\mathbb{R}$ to $\mathbb{R}$ has both a left and a right limit at each point of $\mathbb{R}$, then the set of discontinuities is at most countable.
51. Prove or supply a counter-example: If f is a non-decreasing real valued function on $[0, 1]$ then there is a sequence f_n, $n = 1, 2, \ldots$ of continuous functions on $[0, 1]$ such that for each x in $[0, 1]$, $\lim_{n\to\infty} f_n(x) = f(x)$.
52. Show that if f is a homeomorphism of $[0, 1]$ onto itself then there is a sequence of polynomials $P_n(x)$, $n = 1, 2, \ldots$, such that $P_n \to f$ uniformly on $[0, 1]$ and each P_n is a homeomorphism of $[0, 1]$ onto itself. [Hint: First assume that f is C^1.]
53. Let f be a C^2 function on the real line. Assume that f is bounded with bounded second derivative. Let $A = \sup_x |f(x)|$ and $B = \sup_x |f''(x)|$. Prove that

$$\sup_x |f'(x)| \leq 2\sqrt{AB}.$$

54. Let f be continuous on $\mathbb{R}$ and let

$$f_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} f\left(x + \frac{k}{n}\right).$$

Prove that $f_n(x)$ converges uniformly to a limit on every finite interval $[a, b]$.
55. Let f be a real valued continuous function on the compact interval $[a, b]$. Given $\epsilon > 0$, show that there is a polynomial p such that

$$\begin{aligned} p(a) &= f(a), \\ p'(a) &= 0, \text{ and} \\ |p(x) - f(x)| &< \epsilon, \end{aligned}$$

for all $x \in [a, b]$.
56. A function $f : [0, 1] \to \mathbb{R}$ is said to be **upper semicontinuous** if, given $x \in [0, 1]$ and $\epsilon > 0$, there exists a $\delta > 0$ such that $|y - x| < \delta$ implies that $f(y) < f(x) + \epsilon$. Prove that an upper semicontinuous function on $[0, 1]$ is bounded above and attains its maximum value at some point $p \in [0, 1]$.
57. Let f and f_n be functions from $\mathbb{R}$ to $\mathbb{R}$. Assume that $f_n(x_n) \to f(x)$ as $n \to \infty$ whenever $x_n \to x$. Prove that f is continuous. (Note: the functions f_n are not assumed to be continuous.)

58. Let $f(x)$, $0 \leq x \leq 1$, be a continuous real function with continuous derivative $f'(x)$. Let M be the supremum of $|f'(x)|$, $0 \leq x \leq 1$. Prove: for $n = 1, 2, \ldots$

$$\left|\frac{1}{n}\sum_{k=0}^{n-1} f\left(\frac{k}{n}\right) - \int_0^1 f(x)\,dx\right| \leq \frac{M}{2n}.$$

59. Let K be a compact subset of $\mathbb{R}^m$ and let (B_j) be a sequence of open balls which cover K. Prove that there is an $\epsilon > 0$ such that each ϵ-ball centered at a point of K is contained in at least one of the balls B_j.
60. Let f be a continuous real-valued function on $[0, \infty)$ such that

$$\lim_{x\to\infty}\left(f(x) + \int_0^x f(t)\,dt\right)$$

exists (and is finite). Prove that $\lim_{x\to\infty} f(x) = 0$.
61. A standard theorem asserts that a continuous real-valued function on a compact set is bounded. Prove the converse: if K is a subset of $\mathbb{R}^m$ and if every continuous real-valued function defined on K is bounded, then K is compact.
62. Let $\mathcal{F}$ be a uniformly bounded equicontinuous family of real valued functions defined on the metric space X. Prove that the function

$$g(x) = \sup\{f(x) : f \in \mathcal{F}\}$$

is continuous.
63. Suppose that (f_n) is a sequence of nondecreasing functions which map the unit interval into itself. Suppose that $\lim_{n\to\infty} f_n(x) = f(x)$ pointwise and that f is a continuous function. Prove that $f_n(x) \to f(x)$ uniformly as $n \to \infty$. Note that the functions f_n are not necessarily continuous.
64. Does there exist a continuous real-valued function $f(x)$, $0 \leq x \leq 1$, such that

$$\int_0^1 xf(x)\,dx = 1 \qquad \text{and} \qquad \int_0^1 x^n f(x)\,dx = 0$$

for all $n = 0, 2, 3, 4, 5, \ldots$? Give a proof or counter-example.
65. Let f be a continuous, strictly increasing function from $[0, \infty)$ onto $[0, \infty)$ and let $g = f^{-1}$ (the inverse, not the reciprocal). Prove that

$$\int_0^a f(x)\,dx + \int_0^b g(y)\,dy \geq ab$$

for all positive numbers a, b, and determine the condition for equality.

66. Let f be a function $[0, 1] \to \mathbb{R}$ whose graph $\{(x, f(x)) : x \in [0, 1]\}$ is a closed subset of the unit square. Prove that f is continuous.
67. Let (a_n) be a sequence of positive numbers such that $\sum a_n$ converges. Prove that there exists a sequence of numbers $c_n \to \infty$ as $n \to \infty$ such that $\sum c_n a_n$ converges.
68. Let $f(x, y)$ be a continuous real valued function defined on the unit square $[0, 1] \times [0, 1]$. Prove that $g(x) = \max\{f(x, y) : y \in [0, 1]\}$ is continuous.
69. Let the function f from $[0, 1]$ to $[0, 1]$ have the following properties. It is of class C^1, $f(0) = 0 = f(1)$, and f' is nonincreasing (i.e., f is concave). Prove that the arc-length of the graph of f does not exceed 3.
70. Let A be the set of all positive integers that do not contain the digit 9 in their decimal expansions. Prove that

$$\sum_{a \in A} \frac{1}{a} < \infty.$$

That is, A defines a convergent sub-series of the harmonic series.

5
Multivariable Calculus

This chapter presents the natural geometric theory of calculus in n dimensions.

1 Linear Algebra

It will be taken for granted that you are familiar with the basic concepts of linear algebra — vector spaces, linear transformations, matrices, determinants, and dimension. In particular, you should be aware of the fact that an $m \times n$ matrix A with entries a_{ij} is more than just a static array of mn numbers. It is dynamic. It can act. It defines a **linear transformation** $T_A : \mathbb{R}^n \to \mathbb{R}^m$ that sends n-space to m-space according to the formula

$$T_A(v) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} v_j e_i$$

where $v = \sum v_j e_j \in \mathbb{R}^n$ and $e_1, \ldots, e_n$ is the standard basis of $\mathbb{R}^n$. (Equally, $e_1, \ldots, e_m$ is the standard basis for $\mathbb{R}^m$.)

The set $\mathcal{M} = \mathcal{M}(m, n)$ of all $m \times n$ matrices with real entries a_{ij} is a vector space. Its vectors are matrices. You add two matrices by adding the corresponding entries, $A + B = C$ where $a_{ij} + b_{ij} = c_{ij}$. Similarly, if $\lambda \in \mathbb{R}$ is a scalar, λA is the matrix with entries λa_{ij}. The dimension of the vector space $\mathcal{M}$ is mn, as can be seen by expressing each A as $\sum a_{ij} E_{ij}$ where

E_{ij} is the matrix whose entries are 0, except for the $(ij)^{\text{th}}$ entry which is 1. Thus, as vector spaces, $\mathcal{M} = \mathbb{R}^{mn}$. This gives a natural topology to $\mathcal{M}$.

The set $\mathcal{L} = \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ of linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$ is also a vector space. You combine linear transformations as functions, $U = T + S$ being defined by $U(v) = T(v) + S(v)$, and λT being defined by $(\lambda T)(v) = \lambda T(v)$. The vectors in $\mathcal{L}$ are linear transformations. The mapping $A \mapsto T_A$ is an isomorphism $\mathcal{T} : \mathcal{M} \to \mathcal{L}$. As a rule of thumb, think with linear transformations and compute with matrices.

As explained in Chapter 1, a norm on a vector space V is a function $|\ \ | : V \to \mathbb{R}$ that satisfies three properties:

(a) For all $v \in V$, $|v| \geq 0$; and $|v| = 0$ if and only if $v = 0$.

(b) $|\lambda v| = |\lambda|\,|v|$.

(c) $|v + w| \leq |v| + |w|$.

(Note the abuse of notation in (b); $|\lambda|$ is the magnitude of the scalar λ and $|v|$ is the norm of the vector v.) Norms are used to make vector estimates, and vector estimates underlie multivariable calculus.

A vector space with a norm is a **normed space**. Its norm gives rise to a metric as

$$d(v, v') = \left|v - v'\right|.$$

Thus a normed space is a special kind of metric space.

If V, W are normed spaces, then the **operator norm** of $T : V \to W$ is

$$\|T\| = \sup\{\frac{|Tv|_W}{|v|_V} : v \neq 0\}.$$

The operator norm of T is the **maximum stretch** that T imparts to vectors in V. The subscript on the norm indicates the space in question, which for simplicity is often suppressed.[†]

1 Theorem *Let $T : V \to W$ be a linear transformation from one normed space to another. The following are equivalent:*

(a) $\|T\| < \infty$.

(b) T is uniformly continuous.

(c) T is continuous.

(d) T is continuous at the origin.

Proof Assume (a), $\|T\| < \infty$. For all $v, v' \in V$, linearity of T implies that

$$\left|Tv - Tv'\right| \leq \|T\|\left|v - v'\right|,$$

[†] If $\|T\|$ is finite then T is said to be a **bounded linear transformation**. Unfortunately, this terminology conflicts with T being bounded as a mapping from the metric space V to the metric space W. The only linear transformation that is bounded in the latter sense is the zero transformation.

which gives (b), uniform continuity. Clearly, (b) implies (c) implies (d).

Assume (d) and take $\epsilon = 1$. There is a $\delta > 0$ such that if $u \in V$ and $|u| < \delta$ then

$$|Tu| < 1.$$

For any nonzero $v \in V$, set $u = \lambda v$ where $\lambda = \delta/2\,|v|$. Then $|u| = \delta/2 < \delta$ and

$$\frac{|Tv|}{|v|} = \frac{|Tu|}{|u|} < \frac{1}{|u|} = \frac{2}{\delta}$$

which verifies (a). □

2 Theorem *Any linear transformation $T : \mathbb{R}^n \to W$ is continuous, and if it is an isomorphism then it is a homeomorphism.*

Proof The norm on $\mathbb{R}^n$ is the Euclidean norm

$$|v| = \sqrt{v_1^2 + \cdots + v_n^2}$$

and the norm on W is $|\ \ |_W$. Let $M = \max\{|T(e_1)|_W, \ldots, |T(e_n)|_W\}$. Express $v \in \mathbb{R}^n$ as $v = \sum v_j e_j$. Then $\left|v_j\right| \leq |v|$ and

$$|Tv|_W \leq \sum_{j=1}^{n} \left|T(v_j e_j)\right|_W = \sum_{j=1}^{n} \left|v_j\right| \left|T(e_j)\right|_W \leq n\,|v|\,M$$

which implies that $\|T\| \leq nM < \infty$. By Theorem 1, T is continuous.

Assume that T is an isomorphism. Continuity of T implies that the image of the unit sphere, $T(S^{n-1})$, is a compact subset of W. Injectivity of T implies that the origin of W does not belong to $T(S^{n-1})$. Thus, there is a constant $c > 0$ such that

$$|w| < c \Rightarrow w \notin T(S^{n-1}) \Rightarrow \left|T^{-1}(w)\right| \neq 1.$$

We observe that $\tau = \left|T^{-1}(w)\right| < 1$. For, if not, then $t = 1/\tau < 1$, and we have $\left|T^{-1}(tw)\right| = t\tau = 1$, contrary to the fact that $|tw| < c$. Thus $\left\|T^{-1}\right\| \leq 1/c$ and by Theorem 1, T^{-1} is continuous. A bicontinuous bijection is a homeomorphism. □

3 Corollary *In the world of finite dimensional normed spaces, all linear transformations are continuous and all isomorphisms are homeomorphisms. In particular, $\mathcal{T} : \mathcal{M} \to \mathcal{L}$ is a homeomorphism.*

Proof Let V be an n-dimensional normed space. As you know from linear algebra, there is an isomorphism $H : \mathbb{R}^n \to V$. Any linear transformation $T : V \to W$ factors as

$$T = (T \circ H) \circ H^{-1}.$$

Theorem 2 implies that since $T \circ H$ is a linear transformation defined on $\mathbb{R}^n$, it is continuous, and H is a homeomorphism. Thus T is continuous. If T is an isomorphism, then continuity of T and T^{-1} imply that T is a homeomorphism. □

A fourth norm property involves composites. It states that

(d) $\|T \circ S\| \leq \|T\| \, \|S\|$

for all linear transformations $S : U \to V$ and $T : V \to W$. Thinking in terms of stretch, (d) is clear: S stretches a vector $u \in U$ by at most $\|S\|$, and T stretches $S(u)$ by at most $\|T\|$. The net effect on u is a stretch of at most $\|T\| \, \|S\|$.

Corresponding to composition of linear transformations is the product of matrices. If A is an $m \times k$ matrix and B is a $k \times n$ matrix then the **product matrix** $P = AB$ is the $m \times n$ matrix whose $(ij)^{\text{th}}$ entry is

$$p_{ij} = a_{i1}b_{1j} + \cdots + a_{ik}b_{kj} = \sum_{r=1}^{k} a_{ir}b_{rj}.$$

4 Theorem $T_A \circ T_B = T_{AB}$.

Proof For each $e_r \in \mathbb{R}^k$ and $e_j \in \mathbb{R}^n$ we have

$$T_A(e_r) = \sum_{i=1}^{m} a_{ir}e_i \qquad T_B(e_j) = \sum_{r=1}^{k} b_{rj}e_r.$$

Thus,

$$T_A(T_B(e_j)) = T_A\Big(\sum_{r=1}^{k} b_{rj}e_r\Big) = \sum_{r=1}^{k} b_{rj} \sum_{i=1}^{m} a_{ir}e_i$$

$$= \sum_{i=1}^{m} \sum_{r=1}^{k} a_{ir}b_{rj}e_i = T_{AB}(e_j).$$

Two linear transformations that are equal on a basis are equal. □

Theorem 4 expresses the pleasing fact that matrix multiplication corresponds naturally to composition of linear transformations. See also Exercise 6.

2 Derivatives

A function of a real variable $y = f(x)$ has a derivative $f'(x)$ at x when

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x). \tag{1}$$

If, however, x is a vector variable, (1) makes no sense. For what does it mean to divide by the vector increment h? Equivalent to (1) is the condition

$$f(x+h) = f(x) + f'(x)h + R(h) \quad \Rightarrow \quad \lim_{h \to 0} \frac{R(h)}{|h|} = 0,$$

which is easy to recast in vector terms.

Definition Let $f : U \to \mathbb{R}^m$ be given, where U is an open subset of $\mathbb{R}^n$. The function f is **differentiable** at $p \in U$ with **derivative** $(Df)_p = T$ if $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation and

$$f(p+v) = f(p) + T(v) + R(v) \quad \Rightarrow \quad \lim_{|v| \to 0} \frac{R(v)}{|v|} = 0. \tag{2}$$

We say that the Taylor remainder R is **sublinear** because it tends to 0 faster than $|v|$.

When $n = m = 1$, the multidimensional definition reduces to the standard one. That is because a linear transformation $\mathbb{R} \to \mathbb{R}$ is just multiplication by some real number, in this case multiplication by $f'(x)$.

Here is how to visualize Df. Take $m = n = 2$. The mapping $f : U \to \mathbb{R}^2$ distorts shapes nonlinearly; its derivative describes the linear part of the distortion. Circles are sent by f to wobbly ovals, but they become ellipses under $(Df)_p$. Lines are sent by f to curves, but they become straight lines under $(Df)_p$. See Figure 103 and also Appendix A.

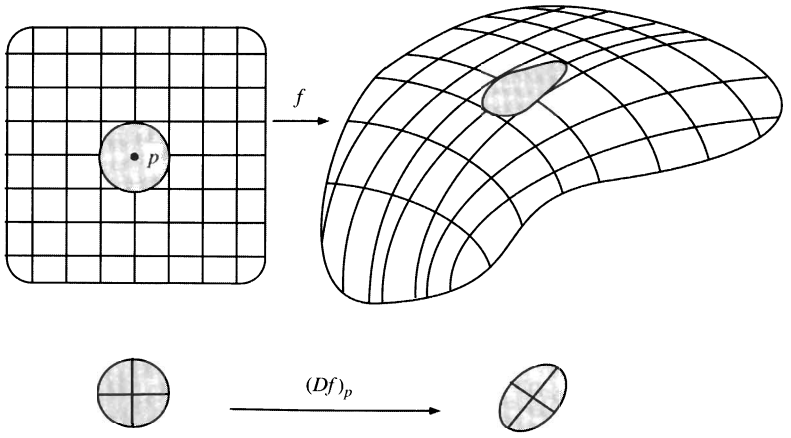


Figure 103 $(Df)_p$ is the linear part of f at p.

This way of looking at differentiability is conceptually simple. Near p, f is the sum of three terms: a constant term $f(p)$, a linear term $(Df)_p v$, and a sublinear remainder term $R(v)$. Keep in mind what kind of an object the derivative is. It is not a number. It is not a vector. No, if it exists, then $(Df)_p$ is a linear transformation from the domain space to the target space.

5 Theorem *If f is differentiable at p, then it unambiguously determines $(Df)_p$ according to the limit formula, valid for all $u \in \mathbb{R}^n$,*

$$(Df)_p(u) = \lim_{t\to 0} \frac{f(p+tu) - f(p)}{t}. \tag{3}$$

Proof Let T be a linear transformation that satisfies (2). Fix any $u \in \mathbb{R}^n$ and take $v = tu$. Then

$$\frac{f(p+tu) - f(p)}{t} = \frac{T(tu) + R(tu)}{t} = T(u) + \frac{R(tu)}{t\,|u|}\,|u|\,.$$

The last term converges to zero as $t \to 0$, which verifies (3). Limits, when they exist, are unambiguous, and therefore if T' is a second linear transformation that satisfies (2) then $T(u) = T'(u)$, so $T = T'$. □

6 Theorem *Differentiability implies continuity.*

Proof Differentiability at p implies that

$$|f(p+v) - f(p)| = \left|(Df)_p v + R(v)\right| \le \left\|(Df)_p\right\| |v| + |R(v)| \to 0$$

as $p + v \to p$. □

Df is the **total derivative** or **Fréchet derivative**. In contrast, the **$(ij)^{\text{th}}$ partial derivative** of f at p is the limit, if it exists,

$$\frac{\partial f_i(p)}{\partial x_j} = \lim_{t\to 0} \frac{f_i(p+te_j) - f_i(p)}{t}.$$

7 Corollary *If the total derivative exists, then the partial derivatives exist, and they are the entries of the matrix that represents the total derivative.*

Proof Substitute in (3) the vector $u = e_j$ and take the i^{th} component of both sides of the resulting equation. □

As is shown in Exercise 15, the mere existence of partial derivatives does not imply differentiability. The simplest sufficient condition beyond the existence of the partials — and the simplest way to recognize differentiability — is given in the next theorem.

8 Theorem *If the partial derivatives of $f : U \to \mathbb{R}^m$ exist and are continuous, then f is differentiable.*

Proof Let A be the matrix of partials at p, $A = [\partial f_i(p)/\partial x_j]$, and let $T : \mathbb{R}^n \to \mathbb{R}^m$ be the linear transformation that A represents. We claim that $(Df)_p = T$. We must show that the Taylor remainder

$$R(v) = f(p+v) - f(p) - Av$$

is sublinear. Draw a path $\sigma = [\sigma_1, \ldots, \sigma_n]$ from p to $q = p + v$ that consists of n segments parallel to the components of v. Thus $v = \sum v_j e_j$ and

$$\sigma_j(t) = p_{j-1} + tv_j e_j \qquad 0 \le t \le 1$$

is a segment from $p_{j-1} = p + \sum_{k<j} v_k e_k$ to $p_j = p_{j-1} + v_j e_j$. See Figure 104.

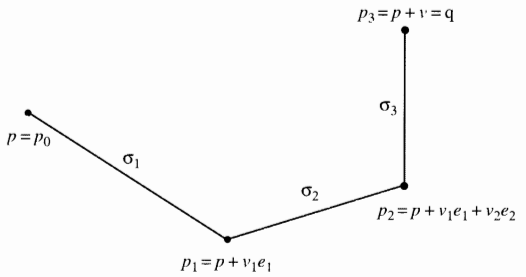


Figure 104 The segmented path σ from p to q.

By the one-dimensional chain rule and mean value theorem applied to the differentiable real-valued function $g(t) = f_i \circ \sigma_j(t)$ of one variable, there exists $t_{ij} \in (0, 1)$ such that

$$f_i(p_j) - f_i(p_{j-1}) = g(1) - g(0) = g'(t_{ij}) = \frac{\partial f_i(p_{ij})}{\partial x_j} v_j,$$

where $p_{ij} = \sigma_j(t_{ij})$. Telescoping $f_i(p+v) - f_i(p)$ along σ gives

$$\begin{aligned} R_i(v) &= f_i(p+v) - f_i(p) - (Av)_i \\ &= \sum_{j=1}^{n} \left(f_i(p_j) - f_i(p_{j-1}) - \frac{\partial f_i(p)}{\partial x_j} v_j \right) \\ &= \sum_{j=1}^{n} \left\{ \frac{\partial f_i(p_{ij})}{\partial x_j} - \frac{\partial f_i(p)}{\partial x_j} \right\} v_j. \end{aligned}$$

Continuity of the partials implies that the terms inside curly brackets tend to 0 as $|v| \to 0$. Thus R is sublinear and f is differentiable at p. □

Next we state and prove the basic rules of multivariable differentiation.

9 Theorem *Let f and g be differentiable. Then*
(a) $D(f+cg) = Df + cDg$.
(b) $D(\text{constant}) = 0$ *and* $D(T(x)) = T$.
(c) $D(g \circ f) = Dg \circ Df$. ***(chain rule)***
(d) $D(f \bullet g) = Df \bullet g + f \bullet Dg$. ***(Leibniz rule)***

There is a fifth rule that concerns the derivative of the nonlinear inversion operator $\text{Inv} : T \mapsto T^{-1}$. It is a glorified version of the formula

$$\frac{d\,x^{-1}}{dx} = -x^{-2},$$

and is discussed in Exercises 32 - 36.

Proof (a) Write the Taylor estimates for f and g and combine them to get the Taylor estimate for $f + cg$.

$$\begin{aligned} f(p+v) &= f(p) + (Df)_p(v) + R_f \\ g(p+v) &= g(p) + (Dg)_p(v) + R_g \\ (f+cg)(p+v) &= (f+cg)(p) + \big((Df)_p + c(Dg)_p\big)(v) + R_f + cR_g. \end{aligned}$$

Since $R_f + cR_g$ is sublinear, $(Df)_p + c(Dg)_p$ is the derivative of $f + cg$ at p.

(b) If $f : \mathbb{R}^n \to \mathbb{R}^m$ is constant, $f(x) = c$ for all $x \in \mathbb{R}^n$, and if $O : \mathbb{R}^n \to \mathbb{R}^m$ denotes the zero transformation then the Taylor remainder $R(v) = f(p+v) - f(p) - O(v)$ is identically zero. Hence $D(\text{ constant })_p = O$.

$T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. If $f(x) = T(x)$, then substituting T itself in the Taylor expression gives the Taylor remainder $R(v) = f(p+v) - f(p) - T(v)$, which is identically zero. Hence $(Df)_p = T$.

Note that when $n = m = 1$, a linear function is of the form $f(x) = ax$, and the previous formula just states that $(ax)' = a$.

(c) Tacitly, we assume that the composite $g \circ f(x) = g(f(x))$ makes sense as x varies in a neighborhood of $p \in U$. The notation $Dg \circ Df$ refers to the composite of linear transformations and is written out as

$$D(g \circ f)_p = (Dg)_q \circ (Df)_p$$

where $q = f(p)$. This chain rule states that the derivative of a composite is the composite of the derivatives. Such a beautiful and natural formula *must* be true. See also Appendix A. Here is a proof:

It is convenient to write the remainder $R(v) = f(p+v) - f(p) - T(v)$ in a different form, defining the scalar function $\boldsymbol{e}(v)$ by

$$\boldsymbol{e}(v) = \begin{cases} \dfrac{|R(v)|}{|v|} & \text{if } v \neq 0 \\ 0 & \text{if } v = 0 \end{cases}$$

Sublinearity is equivalent to $\lim_{v \to 0} \boldsymbol{e}(v) = 0$. Think of $\boldsymbol{e}$ as an **error factor**.

The Taylor expressions for f at p and g at $q = f(p)$ are

$$\begin{aligned} f(p+v) &= f(p) + Av + R_f \\ g(q+w) &= g(q) + Bw + R_g \end{aligned}$$

where $A = (Df)_p$ and $B = (Dg)_q$ as matrices. The composite is expressed as

$$g \circ f(p+v) = g(q + Av + R_f(v)) = g(q) + BAv + BR_f(v) + R_g(w)$$

where $w = Av + R_f(v)$. It remains to show that the remainder terms are sublinear with respect to v. First

$$\big|BR_f(v)\big| \leq \|B\| \,\big|R_f(v)\big|$$

is sublinear. Second,

$$|w| = \big|Av + R_f(v)\big| \leq \|A\|\,|v| + \boldsymbol{e}_f(v)\,|v|\,.$$

Therefore,

$$\big|R_g(w)\big| \leq \boldsymbol{e}_g(w)\,|w| \leq \boldsymbol{e}_g(w)\big(\,\|A\| + \boldsymbol{e}_f(v)\big)\,|v|\,.$$

Since $\boldsymbol{e}_g(w) \to 0$ as $w \to 0$ and since $v \to 0$ implies that w does tend to 0, we see that $R_g(w)$ is sublinear with respect to v. It follows that $(D(g \circ f))_p = BA$ as claimed.

(d) To prove the Leibniz product rule, we must explain the notation $v \bullet w$. In $\mathbb{R}$ there is only one product, the usual multiplication of real numbers. In higher dimensional vector spaces, however, there are many products, and the general way to discuss products is in terms of bilinear maps.

A map $\beta : V \times W \to Z$ is **bilinear** if V, W, Z are vector spaces and for each fixed $v \in V$ the map $\beta(v, .) : W \to Z$ is linear, while for each fixed $w \in W$ the map $\beta(., w) : V \to Z$ is linear. Examples are

(i) Ordinary real multiplication $(x, y) \mapsto xy$ is a bilinear map $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$.
(ii) The dot product is a bilinear map $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.
(iii) The matrix product is a bilinear map $\mathcal{M}(m \times k) \times \mathcal{M}(k \times n) \to \mathcal{M}(m \times n)$.

The precise statement of (d) is that if $\beta : \mathbb{R}^k \times \mathbb{R}^\ell \to \mathbb{R}^m$ is bilinear while $f : U \to \mathbb{R}^k$ and $g : U \to \mathbb{R}^\ell$ are differentiable at p, then the map $x \mapsto \beta(f(x), g(x))$ is differentiable at p and

$$(D\beta(f, g))_p(v) = \beta((Df)_p(v), g(p)) + \beta(f(p), (Dg)_p(v)).$$

Just as a linear transformation between finite-dimensional vector spaces has a finite operator norm, the same is true for bilinear maps:

$$\|\beta\| = \sup\{\frac{|\beta(v, w)|}{|v|\,|w|} : v, w \neq 0\} < \infty.$$

To check this, we view β as a linear map $T_\beta : \mathbb{R}^k \to \mathcal{L}(\mathbb{R}^\ell, \mathbb{R}^m)$. According to Theorems 1, 2, a linear transformation from one finite dimensional normed space to another is continuous and has finite operator norm. Thus the operator norm T_β is finite. That is,

$$\|T_\beta\| = \max\{\frac{\|T_\beta(v)\|}{|v|} : v \neq 0\} < \infty.$$

But $\|T_\beta(v)\| = \max\{|\beta(v, w)| / |w| : w \neq 0\}$, which implies that $\|\beta\| < \infty$.

Returning to the proof of the Leibniz rule, we write out the Taylor estimates for f and g and plug them into β. Using the notation $A = (Df)_p$, $B = (Dg)_p$, bilinearity implies

$$\begin{aligned}\beta(f(p+v), g(p+v)) &= \beta(f(p) + Av + R_f,\ g(p) + Bv + R_g)\\ &= \beta(f(p), g(p)) + \beta(Av, g(p)) + \beta(f(p), Bv)\\ &\quad + \beta(f(p), R_g) + \beta(Av, Bv + R_g) + \beta(R_f, g(p) + Bv + R_g).\end{aligned}$$

The last three terms are sublinear. For

$$\begin{aligned}|\beta(f(p), R_g)| &\leq \|\beta\|\,|f(p)|\,|R_g|\\ |\beta(Av, Bv + R_g)| &\leq \|\beta\|\,\|A\|\,|v|\,|Bv + R_g|\\ |\beta(R_f, g(p) + Bv + R_g| &\leq \|\beta\|\,|R_f|\,|g(p) + Bv + R_g|\end{aligned}$$

Therefore $\beta(f, g)$ is differentiable and $D\beta(f, g) = \beta(Df, g) + \beta(f, Dg)$ as claimed. □

Here are some applications of these differentiation rules:

10 Theorem *A function $f : U \to \mathbb{R}^m$ is differentiable at $p \in U$ if and only if each of its components f_i is differentiable at p. Furthermore, the derivative of its i^{th} component is the i^{th} component of the derivative.*

Proof Assume that f is differentiable at p and express the i^{th} component of f as $f_i = \pi_i \circ f$ where $\pi_i : \mathbb{R}^m \to \mathbb{R}$ is the projection that sends a vector $w = (w_1, \ldots, w_m)$ to w_i. Since π_i is linear it is differentiable. By the chain rule, f_i is differentiable at p and

$$(Df_i)_p = (D\pi_i) \circ (Df)_p = \pi_i \circ (Df)_p.$$

The proof of the converse is equally natural. □

Theorem 10 implies that there is little loss of generality assuming $m = 1$, i.e., that our functions are real-valued. Multidimensionality of the domain, not the target, is what distinguishes multivariable calculus from one-variable calculus.

11 Mean Value Theorem *If $f : U \to \mathbb{R}^m$ is differentiable on U and the segment $[p, q]$ is contained in U, then*

$$|f(q) - f(p)| \leq M\,|q - p|,$$

where $M = \sup\{\|(Df)_x\| : x \in U\}$.

Proof Fix any unit vector $u \in \mathbb{R}^n$. The function

$$g(t) = \langle u,\ f(p + t(q - p))\rangle$$

is differentiable, and we can calculate its derivative. By the one-dimensional Mean Value Theorem, this gives some $\theta \in (0, 1)$ such that $g(1) - g(0) = g'(\theta)$. That is,

$$\langle u,\ f(q) - f(p)\rangle = g'(\theta) = \langle u,\ (Df)_{p+\theta(q-p)}(q - p)\rangle \leq M\,|q - p|.$$

A vector whose dot product with every unit vector is no larger than $M\,|q - p|$ has norm $\leq M\,|q - p|$. □

Remark The one-dimensional Mean Value Theorem is an equality,

$$f(q) - f(p) = f'(\theta)(q - p),$$

and you might expect the same to be true for a vector-valued function if we replace $f'(\theta)$ by $(Df)_\theta$. Not so. See Exercise 17. The closest we can come to an equality form of the multidimensional Mean Value Theorem is the following:

12 C^1 Mean Value Theorem *If $f : U \to \mathbb{R}^m$ is of class C^1 (its derivative exists and is continuous) and if the segment $[p, q]$ is contained in U, then*

$$f(q) - f(p) = T(q - p) \tag{4}$$

*where T is the **average derivative** of f on the segment*

$$T = \int_0^1 (Df)_{p+t(q-p)}\, dt.$$

Conversely, if there is a continuous family of linear maps $T_{pq} \in \mathcal{L}$ for which (4) holds, then f is of class C^1 and $(Df)_p = T_{pp}$.

Proof The integrand takes values in the normed space $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and is a continuous function of t. The integral is the limit of Riemann sums

$$\sum_k (Df)_{p+t_k(q-p)}\, \Delta t_k,$$

which lie in $\mathcal{L}$. Since the integral is an element of $\mathcal{L}$, it has a right to act on the vector $q - p$. Alternately, if you integrate each entry of the matrix that represents Df along the segment, the resulting matrix represents T. Fix an index i, and apply the Fundamental Theorem of Calculus to the C^1 real-valued function of one variable

$$g(t) = f_i \circ \sigma(t),$$

where $\sigma(t) = p + t(q - p)$ parameterizes $[p, q]$. This gives

$$\begin{aligned} f_i(q) - f_i(p) = g(1) - g(0) &= \int_0^1 g'(t)\, dt \\ &= \int_0^1 \sum_{j=1}^n \frac{\partial f_i(\sigma(t))}{\partial x_j}(q_j - p_j)\, dt \\ &= \sum_{j=1}^n \int_0^1 \frac{\partial f_i(\sigma(t))}{\partial x_j}\, dt\, (q_j - p_j), \end{aligned}$$

which is the i^{th} component of $T(q - p)$.

To check the converse, we assume that (4) holds for a continuous family of linear maps T_{pq}. Take $q = p + v$. The first-order Taylor remainder at p is

$$R(v) = f(p + v) - f(p) - T_{pp}(v) = \big(T_{pq} - T_{pp}\big)(v),$$

which is sublinear with respect to v. Therefore $(Df)_p = T_{pp}$. □

13 Corollary *Assume that U is connected. If $f : U \to \mathbb{R}^m$ is differentiable and for each point $x \in U$, $(Df)_x = 0$, then f is constant*

Proof The enjoyable open and closed argument is left to you as Exercise 20. □

We conclude this section with another useful rule — differentiation past the integral. See also Exercise 23.

14 Theorem *Assume that $f : [a, b] \times (c, d) \to \mathbb{R}$ is continuous and that $\partial f(x, y)/\partial y$ exists and is continuous. Then*

$$F(y) = \int_a^b f(x, y)\, dx$$

is of class C^1 and

$$\frac{dF}{dy} = \int_a^b \frac{\partial f(x, y)}{\partial y}\, dx. \tag{5}$$

Proof By the C^1 Mean Value Theorem, if h is small, then

$$\frac{F(y + h) - F(y)}{h} = \frac{1}{h} \int_a^b \left(\int_0^1 \frac{\partial f(x, y + th)}{\partial y}\, dt \right) h\, dx.$$

The inner integral is the average partial derivative of f with respect to y along the segment from y to $y + h$. Continuity implies that this average converges to $\partial f(x, y)/\partial y$ as $h \to 0$, which verifies (5). Continuity of dF/dy follows from continuity of $\partial f/\partial y$. See Exercise 22. □

3 Higher derivatives

In this section we define higher-order multivariable derivatives. We do so in the same spirit as in the previous section: the second derivative will be the derivative of the first derivative, viewed naturally. Assume that $f : U \to \mathbb{R}^m$ is differentiable on U. The derivative $(Df)_x$ exists at each $x \in U$ and the map $x \mapsto (Df)_x$ defines a function

$$Df : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m).$$

The derivative Df is the same sort of thing that f is, namely a function from an open subset of a vector space into another vector space. In the case

of Df, the target vector space is not $\mathbb{R}^m$, but rather the mn dimensional space $\mathcal{L}$. If Df is differentiable at $p \in U$ then by definition

$$(D(Df))_p = (D^2 f)_p = \text{ the } \textbf{second derivative} \text{ of } f \text{ at } p,$$

and f is **second-differentiable** at p. The second derivative is a linear map from $\mathbb{R}^n$ into $\mathcal{L}$. For each $v \in \mathbb{R}^n$, $(D^2 f)_p(v)$ belongs to $\mathcal{L}$ and therefore is a linear transformation $\mathbb{R}^n \to \mathbb{R}^m$, so $(D^2 f)_p(v)(w)$ is bilinear, and we write it as

$$(D^2 f)_p(v, w).$$

(Recall that bilinearity is linearity in each variable separately.)

Third- and higher derivatives are defined in the same way. If f is second differentiable on U, then $x \mapsto (D^2 f)_x$ defines a map

$$D^2 f : U \to \mathcal{L}^2$$

where $\mathcal{L}^2$ is the vector space of bilinear maps $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$. If $D^2 f$ is differentiable at p, then f is third differentiable there and its third derivative is the trilinear map $(D^3 f)_p = (D(D^2 f))_p$.

Just as for first derivatives, the relation between the second derivative and the second partial derivatives calls for thought. Express $f : U \to \mathbb{R}^m$ in component form as $f(x) = (f_1(x), \ldots, f_m(x))$ where x varies in U.

15 Theorem *If $(D^2 f)_p$ exists then $(D^2 f_k)_p$ exists, the second-partials at p exist, and*

$$(D^2 f_k)_p(e_i, e_j) = \frac{\partial^2 f_k(p)}{\partial x_i \partial x_j}.$$

Conversely, existence of the second-partials implies existence of $(D^2 f)_p$, provided that the second-partials exist at all points $x \in U$ near p and are continuous at p.

Proof Assume that $(D^2 f)_p$ exists. Then $x \mapsto (Df)_x$ is differentiable at $x = p$ and the same is true of the matrix

$$M_x = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}$$

that represents it; $x \mapsto M_x$ is differentiable at $x = p$. For, according to Theorem 10, a vector function is differentiable if and only if its components

are differentiable; and then, the derivative of the k^{th} component is the k^{th} component of the derivative. A matrix is a special type of vector, its components are its entries. Thus the entries of M_x are differentiable at $x = p$, and the second-partials exist. Furthermore, the k^{th} row of M_x is a differentiable vector function of x at $x = p$ and

$$(D(Df_k))_p(e_i)(e_j) = (D^2 f_k)_p(e_i, e_j) = \lim_{t \to 0} \frac{(Df_k)_{p+te_i}(e_j) - (Df_k)_p(e_j)}{t}.$$

The first derivatives appearing in this fraction are the j^{th} partials of f_k at $p + te_i$ and at p. Thus, $\partial^2 f_k(p)/\partial x_i \partial x_j = (D^2 f_k)_p(e_i, e_j)$ as claimed.

Conversely, assume that the second-partials exist at all x near p and are continuous at p. Then the entries of M_x have partials that exist at all points q near p, and are continuous at p. Theorem 8 implies that $x \mapsto M_x$ is differentiable at $x = p$; i.e., f is second-differentiable at p. □

The most important and surprising property of second derivatives is symmetry.

16 Theorem *If $(D^2 f)_p$ exists then it is symmetric: for all $v, w \in \mathbb{R}^n$,*

$$(D^2 f)_p(v, w) = (D^2 f)_p(w, v).$$

Proof We will assume that f is real-valued (i.e., $m = 1$) because the symmetry assertion concerns the arguments of f, not its values. For a variable $t \in [0, 1]$, draw the parallelogram P determined by the vectors tv, tw, and label the vertices with ± 1's as in Figure 105.

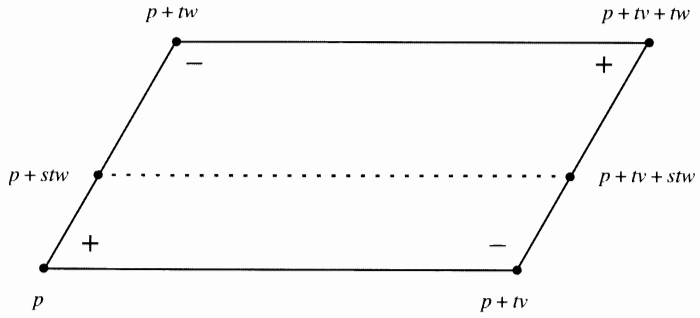


Figure 105 The parallelogram P has signed vertices.

The quantity

$$\Delta = \Delta(t, v, w) = f(p + tv + tw) - f(p + tv) - f(p + tw) + f(p)$$

is the signed sum of f at the vertices of P. Clearly, Δ is symmetric with respect to v, w,

$$\Delta(t, v, w) = \Delta(t, w, v).$$

We claim that

$$(D^2 f)_p(v, w) = \lim_{t \to 0} \frac{\Delta(t, v, w)}{t^2}, \tag{6}$$

from which symmetry of $D^2 f$ follows.

Fix t, v, w and write $\Delta = g(1) - g(0)$ where

$$g(s) = f(p + tv + stw) - f(p + stw).$$

Since f is differentiable, so is g. By the one-dimensional Mean Value Theorem there exists $\theta \in (0, 1)$ with $\Delta = g'(\theta)$. By the Chain Rule, $g'(\theta)$ can be written in terms of Df and we get

$$\Delta = g'(\theta) = (Df)_{p+tv+\theta tw}(tw) - (Df)_{p+\theta tw}(tw).$$

Taylor's estimate applied to the differentiable function $u \mapsto (Df)_u$ at $u = p$ gives

$$(Df)_{p+x} = (Df)_p + (D^2 f)_p(x, \ .\ ) + R(x, \ .\ )$$

where $R(x, \ .\ ) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is sublinear with respect to x. Writing out this estimate for $(Df)_{p+x}$ first with $x = tv + \theta tw$ and then with $x = \theta tw$ gives

$$\begin{aligned}\frac{\Delta}{t^2} = \frac{1}{t}\Big\{&\Big[(Df)_p(w) + (D^2 f)_p(tv + \theta tw, w) + R(tv + \theta tw, w)\Big] \\ &- \Big[(Df)_p(w) + (D^2 f)_p(\theta tw, w) + R(\theta tw, w)\Big]\Big\} \\ = (D^2 f)_p(v, w) &+ \frac{R(tv + \theta tw, w)}{t} - \frac{R(\theta tw, w)}{t}\end{aligned}$$

Bilinearity was used to combine the two second derivative terms. Sublinearity of $R(x, w)$ with respect to x implies that the last two terms tend to 0 as $t \to 0$, which completes the proof of (6). Since $(D^2 f)_p$ is the limit of a symmetric (although nonlinear) function of v, w it too is symmetric. □

Remark. The fact that $D^2 f$ can be expressed directly as a limit of values of f is itself interesting. It should remind you of its one-dimensional counterpart,

$$f''(x) = \lim_{h \to 0} \frac{f(x + h) + f(x - h) - 2f(x)}{h^2}.$$

17 Corollary *Corresponding mixed second-partials of a second-differentiable function are equal,*

$$\frac{\partial^2 f_k(p)}{\partial x_i \partial x_j} = \frac{\partial^2 f_k(p)}{\partial x_j \partial x_i}.$$

Proof The equalities

$$\frac{\partial^2 f_k(p)}{\partial x_i \partial x_j} = (D^2 f_k)_p(e_i, e_j) = (D^2 f_k)_p(e_j, e_i) = \frac{\partial^2 f_k(p)}{\partial x_j \partial x_i}$$

follow from Theorem 15 and the symmetry of $D^2 f$. □

The mere existence of the second-partials does not imply second-differentiability, nor does it imply equality of corresponding mixed second-partials. See Exercise 24.

18 Corollary *The r^{th} derivative, if it exists, is symmetric: permutation of the vectors $v_1, \ldots, v_r$ does not affect the value of $(D^r f)_p(v_1, \ldots, v_r)$. Corresponding mixed higher-order partials are equal.*

Proof The induction argument is left to you as Exercise 29. □

In my opinion Theorem 16 is quite natural, even though its proof is tricky. It proceeds from a pointwise hypothesis to a pointwise conclusion: whenever the second derivative exists it is symmetric. No assumption is made about continuity of partials. It is possible that f is second-differentiable at p and nowhere else. See Exercise 25. All the same, it remains standard to prove equality of mixed partials under stronger hypotheses, namely that $D^2 f$ is continuous. See Exercise 27.

We conclude this section with a brief discussion of the rules of higher-order differentiation. It is simple to check that the r^{th} derivative of $f + cg$ is $D^r f + cD^r g$. Also, if β is k-linear and $k < r$ then $f(x) = \beta(x, \ldots, x)$ has $D^r f = 0$. On the other hand, if $k = r$ then $(D^r f)_p = r!\,\mathrm{Symm}(\beta)$, the where $\mathrm{Symm}(\beta)$ is the symmetrization of β. See Exercise 28.

The chain rule for r^{th} derivatives is a bit complicated. The difficulties arise from the fact that x appears in two places in the expression for the first order chain rule, $(D\,g \circ f)_x = (Dg)_{f(x)} \circ (Df)_x$, and so differentiating this product produces

$$(D^2 g)_{f(x)} \circ (Df)_x^2 \quad + \quad (Dg)_{f(x)} \circ (D^2 f)_x.$$

(The meaning of $(Df)_x^2$ needs clarification.) Differentiating again produces four terms, two of which combine. The general formula is

$$(D^r g \circ f)_x = \sum_{k=1}^{r} \sum_{\mu} (D^k g)_{f(x)} \circ (D^\mu f)_x$$

where the sum on μ is taken as μ runs through all partitions of $\{1, \ldots, r\}$ into k disjoint subsets. See Exercise 41.

The higher-order Leibniz rule is left for you as Exercise 42.

4 Smoothness Classes

A map $f : U \to \mathbb{R}^m$ is of **class C^r** if it is r^{th} order differentiable at each $p \in U$ and its derivatives depend continuously on p. (Since differentiability implies continuity, all the derivatives of order $< r$ are automatically continuous; only the r^{th} derivative is in question.) If f is of class C^r for all r, it is **smooth** or of **class C^∞**. According to the differentiation rules, these smoothness classes are closed under the operations of linear combination, product, and composition. We discuss next how they are closed under limits.

Let (f_k) be a sequence of C^r functions $f_k : U \to \mathbb{R}^m$. The sequence is

(a) **Uniformly C^r convergent** if for some C^r function $f : U \to \mathbb{R}^m$,

$$\begin{aligned} f_k &\rightrightarrows f \\ Df_k &\rightrightarrows Df \\ &\cdots \\ D^r f_k &\rightrightarrows D^r f \end{aligned}$$

as $k \to \infty$.

(b) **Uniformly C^r Cauchy** if for each $\epsilon > 0$ there is an N such that for all $k, \ell \geq N$ and all $x \in U$,

$$\begin{aligned} |f_k(x) - f_\ell(x)| &< \epsilon \\ \|(Df_k)_x - (Df_\ell)_x\| &< \epsilon \\ &\cdots \\ \left\|(D^r f_k)_x - (D^r f_\ell)_x\right\| &< \epsilon. \end{aligned}$$

19 Theorem *Uniform C^r convergence and Cauchyness are equivalent.*

Proof Convergence always implies the Cauchy condition. As for the converse, first assume that $r = 1$. We know that f_k converges uniformly to a continuous function f, and the derivative sequence converges uniformly to a continuous limit

$$Df_k \rightrightarrows G.$$

We claim that $Df = G$. Fix $p \in U$ and consider points q in a small convex neighborhood of p. The C^1 Mean Value Theorem and uniform convergence imply that as $k \to \infty$,

$$\begin{aligned} f_k(q) - f_k(p) &= \int_0^1 (Df_k)_{p+t(q-p)}\, dt\, (q-p) \\ \downarrow\downarrow & \qquad\qquad \downarrow\downarrow \\ f(q) - f(p) &= \int_0^1 G(p + t(q-p))\, dt\, (q-p). \end{aligned}$$

This integral of G is a continuous function of q that reduces to $G(p)$ when $p = q$. By the converse part of the C^1 Mean Value Theorem, f is differentiable and $Df = G$. Therefore f is C^1 and f_k converges C^1 uniformly to f as $k \to \infty$, completing the proof when $r = 1$.

Now suppose that $r \geq 2$. The maps $Df_k : U \to \mathcal{L}$ form a uniformly C^{r-1} Cauchy sequence. The limit, by induction, is C^{r-1} uniform; i.e., as $k \to \infty$,

$$D^s(Df_k) \rightrightarrows D^s G$$

for all $s \leq r - 1$. Hence f_k converges C^r uniformly to f as $k \to \infty$, completing the induction. □

The **C^r norm** of a C^r function $f : U \to \mathbb{R}^m$ is

$$\|f\|_r = \max\{\sup_{x \in U} |f(x)|, \ldots, \sup_{x \in U} \left\|(D^r f)_x\right\|\}.$$

The set of functions with $\|f\|_r < \infty$ is denoted $C^r(U, \mathbb{R}^m)$.

20 Corollary $\| \ \|_r$ *makes $C^r(U, \mathbb{R}^m)$ a* ***Banach space*** *— a complete normed vector space.*

Proof The norm properties are easy to check; completeness follows from Theorem 19. □

21 C^r *M*-test *If $\sum M_k$ is a convergent series of constants and if $\|f_k\|_r \leq M_k$ for all k, then the series of functions $\sum f_k$ converges in $C^r(U, \mathbb{R}^m)$ to a function f. Term by term differentiation of order $\leq r$ is valid, $D^r f = \sum_k D^r f_k$.*

Proof Obvious from the preceding corollary. □

5 Implicit and Inverse Functions

Let $f : U \to \mathbb{R}^m$ be given, where U is an open subset of $\mathbb{R}^n \times \mathbb{R}^m$. Fix attention on a point $(x_0, y_0) \in U$ and write $f(x_0, y_0) = z_0$. Our goal is to solve the equation

$$f(x, y) = z_0 \tag{7}$$

near (x_0, y_0). More precisely, we hope to show that the set of points (x, y) nearby (x_0, y_0) at which $f(x, y) = z_0$, the so-called **z_0-locus** of f, is the graph of a function $y = g(x)$. If so, g is the **implicit function** defined by (7). See Figure 106.

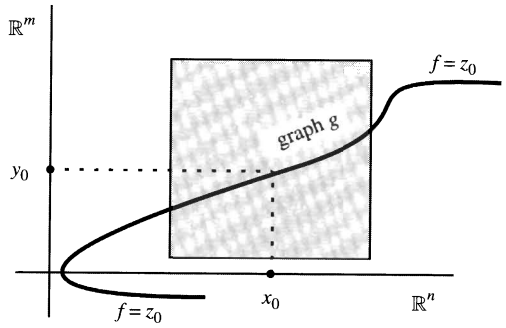


Figure 106 Near (x_0, y_0) the z_0-locus of f is the graph of a function $y = g(x)$.

Under various hypotheses we will show that g exists, is unique, and is differentiable. The main assumption, which we make throughout this section, is that

$$\text{the } m \times m \text{ matrix} \quad B = \left[\frac{\partial f_i(x_0, y_0)}{\partial y_j}\right] \quad \text{is invertible.}$$

Equivalently the linear transformation that B represents is an isomorphism $\mathbb{R}^m \to \mathbb{R}^m$.

22 Implicit Function Theorem *If the function f above is C^r, $1 \le r \le \infty$, then near (x_0, y_0), the z_0-locus of f is the graph of a unique function $y = g(x)$. Besides, g is C^r.*

Proof Without loss of generality, we suppose that (x_0, y_0) is the origin $(0, 0)$ in $\mathbb{R}^n \times \mathbb{R}^m$, and $z_0 = 0$ in $\mathbb{R}^m$. The Taylor expression for f is

$$f(x, y) = Ax + By + R$$

where A is the $m \times n$ matrix

$$A = \left[\frac{\partial f_i(x_0, y_0)}{\partial x_j}\right]$$

and R is sublinear. Then, solving $f(x, y) = 0$ for $y = gx$ is equivalent to solving

$$y = -B^{-1}(Ax + R(x, y)). \tag{8}$$

In the unlikely event that R does not depend on y, (8) is an explicit formula for gx and the implicit function is an explicit function. In general, the idea is that the remainder R depends so weakly on y that we can switch it to the left hand side of (8), absorbing it in the y term.

Solving (8) for y as a function of x is the same as finding a fixed point of

$$K_x : y \mapsto -B^{-1}(Ax + R(x, y)),$$

so we hope to show that K_x contracts. The remainder R is a C^1 function, and $(DR)_{(0,0)} = 0$. Therefore if r is small and $|x|, |y| \le r$ then

$$\|B^{-1}\| \left\|\frac{\partial R(x, y)}{\partial y}\right\| \le \frac{1}{2}.$$

By the Mean Value Theorem this implies that

$$\begin{aligned} |K_x(y_1) - K_x(y_2)| &\le \|B^{-1}\| \, |R(x, y_1) - R(x, y_2)| \\ &\le \|B^{-1}\| \left\|\frac{\partial R}{\partial y}\right\| |y_1 - y_2| \le \frac{1}{2}|y_1 - y_2| \end{aligned}$$

for $|x|, |y_1|, |y_2| \le r$. Due to continuity at the origin, if $|x| \le \tau \ll r$ then

$$|K_x(0)| \le \frac{1}{2}.$$

Thus, for all $x \in X$, K_x contracts Y into itself where X is the τ-neighborhood of 0 in $\mathbb{R}^n$ and Y is the closure of the r-neighborhood of 0 in $\mathbb{R}^m$. See Figure 107.

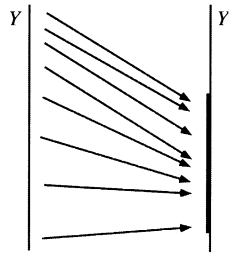


Figure 107 K_x contracts Y into itself.

By the Contraction Mapping Principle, K_x has a unique fixed point $g(x)$ in Y. This implies that near the origin, the zero locus of f is the graph of a function $y = g(x)$.

It remains to check that g is C^r. First we show that g obeys a Lipschitz condition at 0. We have

$$\begin{aligned} |gx| &= |K_x(gx) - K_x(0) + K_x(0)| \le \mathrm{Lip}(K_x)\,|gx - 0| + |K_x(0)| \\ &\le \frac{1}{2}\,|gx| + \left|B^{-1}(Ax + R(x,0))\right| \le \frac{1}{2}\,|gx| + 2L\,|x| \end{aligned}$$

where $L = \left\|B^{-1}\right\| \|A\|$ and $|x|$ is small. Thus g satisfies the Lipschitz condition

$$|gx| \le 4L\,|x|\,.$$

In particular, g is continuous at $x = 0$.

Note the trick here. The term $|gx|$ appears on both sides of the inequality but since its coefficient on the r.h.s. is smaller than that on the l.h.s., they combine to give a nontrivial inequality.

By the chain rule, the derivative of g at the origin, if it does exist, must satisfy $A + B(Dg)_0 = 0$, so we aim to show that $(Dg)_0 = -B^{-1}A$. Since gx is a fixed point of K_x, we have $gx = -B^{-1}A(x + R)$ and the Taylor estimate for g at the origin is

$$\begin{aligned} \left|g(x) - g(0) - (-B^{-1}Ax)\right| &= \left|B^{-1}R(x,gx)\right| \le \left\|B^{-1}\right\| |R(x,gx)| \\ &\le \left\|B^{-1}\right\| e(x,gx)(|x| + |gx|) \\ &\le \left\|B^{-1}\right\| e(x,gx)(1 + 4L)\,|x| \end{aligned}$$

where $e(x, y) \to 0$ as $(x, y) \to (0, 0)$. Since $gx \to 0$ as $x \to 0$, the error factor $e(x, gx)$ does tend to 0 as $x \to 0$, the remainder is sublinear with respect to x, and g is differentiable at 0 with $(Dg)_0 = -B^{-1}A$.

All facts proved at the origin hold equally at points (x, y) on the zero locus near the origin. For the origin is nothing special. Thus, g is differentiable at x and $(Dg)_x = -B_x^{-1} \circ A_x$ where

$$A_x = \frac{\partial f(x, gx)}{\partial x} \qquad B_x = \frac{\partial f(x, gx)}{\partial y}.$$

Since gx is continuous (being differentiable) and f is C^1, A_x and B_x are continuous functions of x. According to Cramer's Rule for finding the inverse of a matrix, the entries of B_x^{-1} are explicit, algebraic functions of the entries of B_x, and therefore they depend continuously on x. Therefore g is C^1.

To complete the proof that g is C^r, we apply induction. For $2 \le r < \infty$, assume the theorem is true for $r - 1$. When f is C^r this implies that g is

C^{r-1}. Because they are composites of C^{r-1} functions, A_x and B_x are C^{r-1}. Because the entries of B_x^{-1} depend algebraically on the entries of B_x, B_x^{-1} is also C^{r-1}. Therefore $(Dg)_x$ is C^{r-1} and g is C^r. If f is C^∞, we have just shown that g is C^r for all finite r, and thus g is C^∞. □

Exercises 35 and 36 discuss the properties of matrix inversion avoiding Cramer's Rule and finite dimensionality.

Next we are going to deduce the Inverse Function Theorem from the Implicit Function Theorem. A fair question is: since they turn out to be equivalent theorems, why not do it the other way around? Well, in my own experience, the Implicit Function Theorem is more basic and flexible. I have at times needed forms of the Implicit Function Theorem with weaker differentiability hypotheses respecting x than y and they do not follow from the Inverse Function Theorem. For example, if we merely assume that $B = \partial f(x_0, y_0)/\partial y$ is invertible, that $\partial f(x, y)/\partial x$ is a continuous function of (x, y), and that f is continuous (or Lipschitz) then the local implicit function of f is continuous (or Lipschitz). It is not necessary to assume that f is of class C^1.

Just as a homeomorphism is a continuous bijection whose inverse is continuous, so a ***C^r* diffeomorphism** is a C^r bijection whose inverse is C^r. (We assume $1 \le r \le \infty$.) The inverse being C^r is not automatic. The example to remember is $f(x) = x^3$. It is a C^∞ bijection $\mathbb{R} \to \mathbb{R}$ and is a homeomorphism but not a diffeomorphism because its inverse fails to be differentiable at the origin. Since differentiability implies continuity, every diffeomorphism is a homeomorphism.

Diffeomorphisms are to C^r things as isomorphisms are to algebraic things. The sphere and ellipsoid are diffeomorphic under a diffeomorphism $\mathbb{R}^3 \to \mathbb{R}^3$, but the sphere and the surface of the cube are only homeomorphic, not diffeomorphic.

23 Inverse Function Theorem *If $m = n$ and $f : U \to \mathbb{R}^m$ is C^r, $1 \le r \le \infty$, and if at some $p \in U$, $(Df)_p$ is an isomorphism, then f is a C^r diffeomorphism from a neighborhood of p to a neighborhood of $f(p)$.*

Proof Set $F(x, y) = f(x) - y$. Clearly F is C^r, $F(p, fp) = 0$, and the derivative of F with respect to x at (p, fp) is $(Df)_p$. Since $(Df)_p$ is an isomorphism, we can apply the implicit function theorem (with x and y interchanged!) to find neighborhoods U_0 of p and V of fp and a C^r implicit function $h : V \to U_0$ uniquely defined by the equation

$$F(hy, y) = 0.$$

Then $f(hy) = y$, so $f \circ h = \mathrm{id}_V$ and h is a right inverse of f.

Except for a little fussy set theory, this completes the proof: f bijects $U_1 = \{x \in U_0 : fx \in V\}$ onto V and its inverse is h, which we know to be a C^r map. To be precise, we must check three things,

(a) U_1 is a neighborhood of p.

(b) h is a right inverse of $f\big|_{U_1}$. That is, $f\big|_{U_1} \circ h = \mathrm{id}_V$.

(c) h is a left inverse of $f\big|_{U_1}$. That is, $h \circ f\big|_{U_1} = \mathrm{id}_{U_1}$.

See Figure 108.

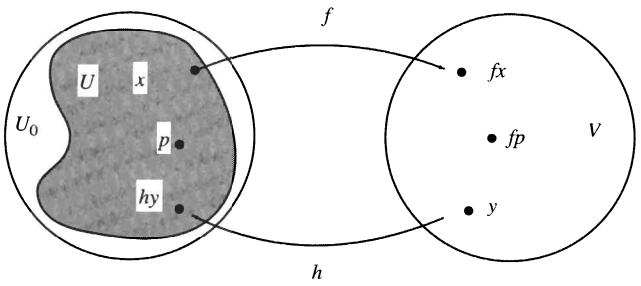


Figure 108 f is a local diffeomorphism.

(a) Since f is continuous, U_1 is open. Since $p \in U_0$ and $fp \in V$, p belongs to U_1.

(b) Take any $y \in V$. Since $hy \in U_0$ and $f(hy) = y$, we see that $hy \in U_1$. Thus, $f\big|_{U_1} \circ h$ is well defined and $f\big|_{U_1} \circ h(y) = f \circ h(y) = y$.

(c) Take any $x \in U_1$. By definition of U_1, $fx \in V$ and there is a unique point $h(fx)$ in U_0 such that $F(h(fx), fx) = 0$. Observe that x itself is just such a point. It lies in U_0 because it lies in U_1, and it satisfies $F(x, fx) = 0$ since $F(x, fx) = fx - fx$. By uniqueness of h, $h(f(x)) = x$. □

Upshot If $(Df)_p$ is an isomorphism, then f is a local diffeomorphism at p.

6* The Rank Theorem

The **rank** of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is the dimension of its range. In terms of matrices, the rank is the size of the largest minor with nonzero determinant. If T is onto then its rank is m. If it is one-to-one, its rank is n. A standard formula in linear algebra states that

$$\text{rank } T + \text{nullity } T = n$$

where nullity is the dimension of the kernel of T. A differentiable function $f : U \to \mathbb{R}^m$ has **constant rank** k if for all $p \in U$ the rank of $(Df)_p$ is k.

An important property of rank is that if T has rank k and $\|S - T\|$ is small, then S has rank $\geq k$. The rank of T can increase under a small perturbation of T but it cannot decrease. Thus, if f is C^1 and $(Df)_p$ has rank k then automatically $(Df)_x$ has rank $\geq k$ for all x near p. See Exercise 43.

The Rank Theorem describes maps of constant rank. It says that locally they are just like linear projections. To formalize this we say that maps $f : A \to B$ and $g : C \to D$ are equivalent (for want of a better word) if there are bijections $\alpha : A \to C$ and $\beta : B \to D$ such that $g = \beta \circ f \circ \alpha^{-1}$. An elegant way to express this equation is as commutativity of the diagram

$$\begin{array}{ccc} A & \xrightarrow{\ f\ } & B \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle\beta} \\ C & \xrightarrow{\ g\ } & D\,. \end{array}$$

Commutativity means that for each $a \in A$, $\beta(f(a)) = g(\alpha(a))$. Following the maps around the rectangle clockwise from A to D gives the same result as following them around it counterclockwise. The α, β are "changes of variable." If f, g are C^r and α, β are C^r diffeomorphisms, $1 \leq r \leq \infty$, then f and g are said to be $\boldsymbol{C^r}$ **equivalent**, and we write $f \approx_r g$. As C^r maps, f and g are indistinguishable.

24 Lemma *C^r equivalence is an equivalence relation and it has no effect on rank.*

Proof Since diffeomorphisms form a group, $\approx_r$ is an equivalence relation. Also, if $g = \beta \circ f \circ \alpha^{-1}$, then the chain rule implies that

$$Dg = D\beta \circ Df \circ D\alpha^{-1}.$$

Since $D\beta$ and $D\alpha^{-1}$ are isomorphisms, Df and Dg have equal rank. □

The linear projection $P : \mathbb{R}^n \to \mathbb{R}^m$

$$P(x_1, \ldots, x_n) = (x_1, \ldots, x_k, 0, \ldots, 0)$$

has rank k. It projects $\mathbb{R}^n$ onto the k-dimensional subspace $\mathbb{R}^k \times 0$. The matrix of P is

$$\begin{bmatrix} I_{k\times k} & 0 \\ 0 & 0 \end{bmatrix}.$$

25 Rank Theorem *Locally, a C^r constant rank k map is C^r equivalent to a linear projection onto a k-dimensional subspace.*

As an example, think of the radial projection $\pi : \mathbb{R}^3 \setminus \{0\} \to S^2$, where $\pi(v) = v/|v|$. It has constant rank 2, and is locally indistinguishable from linear projection of $\mathbb{R}^3$ to the (x, y)-plane.

Proof Let $f : U \to \mathbb{R}^m$ have constant rank k and let $p \in U$ be given. We will show that on a neighborhood of p, $f \approx_r P$.

Step 1. Define translations of $\mathbb{R}^n$ and $\mathbb{R}^m$ by

$$\begin{aligned} \tau : \mathbb{R}^n &\to \mathbb{R}^n & \tau' : \mathbb{R}^m &\to \mathbb{R}^m \\ z &\mapsto z + p & z' &\mapsto z' - fp \end{aligned}$$

The translations are diffeomorphisms of $\mathbb{R}^n$ and $\mathbb{R}^m$ and they show that f is C^r equivalent to $\tau' \circ f \circ \tau$, a C^r map that sends 0 to 0 and has constant rank k. Thus, it is no loss of generality to assume in the first place that p is the origin in $\mathbb{R}^n$ and fp is the origin in $\mathbb{R}^m$. We do so.

Step 2. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an isomorphism that sends $0 \times \mathbb{R}^{n-k}$ onto the kernel of $(Df)_0$. Since the kernel has dimension $n-k$, there is such a T. Let $T' : \mathbb{R}^m \to \mathbb{R}^m$ be an isomorphism that sends the image of $(Df)_0$ onto $\mathbb{R}^k \times 0$. Since $(Df)_0$ has rank k there is such a T'. Then $f \approx_r T' \circ f \circ T$. This map sends the origin in $\mathbb{R}^n$ to the origin in $\mathbb{R}^m$, its derivative at the origin has kernel $0 \times \mathbb{R}^{n-k}$, and its image $\mathbb{R}^k \times 0$. Thus, it is no loss of generality to assume in the first place that f has these properties. We do so.

Step 3. Write

$$(x, y) \in \mathbb{R}^k \times \mathbb{R}^{n-k} \qquad f(x, y) = (f_X(x, y),\ f_Y(x, y)) \in \mathbb{R}^k \times \mathbb{R}^{m-k}.$$

We are going to find a $g \approx_r f$ such that

$$g(x, 0) = (x, 0).$$

The matrix of $(Df)_0$ is

$$\begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix}$$

where A is $k \times k$ and invertible. Thus, by the Inverse Function Theorem, the map

$$\sigma : x \mapsto f_X(x, 0)$$

is a diffeomorphism $\sigma : X \to X'$ where X, X' are small neighborhoods of the origin in $\mathbb{R}^k$. For $x' \in X'$, set

$$h(x') = f_Y(\sigma^{-1}(x'), 0).$$

This makes h a C^r map $X' \to \mathbb{R}^{m-k}$, and

$$h(\sigma(x)) = f_Y(x, 0).$$

The image of $X \times 0$ under f is the graph of h. For

$$\begin{aligned} f(X \times 0) &= \{f(x, 0) : x \in X\} = \{(f_X(x, 0),\ f_Y(x, 0)) : x \in X\} \\ &= \{(f_X(\sigma^{-1}(x'), 0),\ f_Y(\sigma^{-1}(x'), 0)) : x' \in X'\} \\ &= \{(x',\ h(x')) : x' \in X'\}. \end{aligned}$$

See Figure 109.

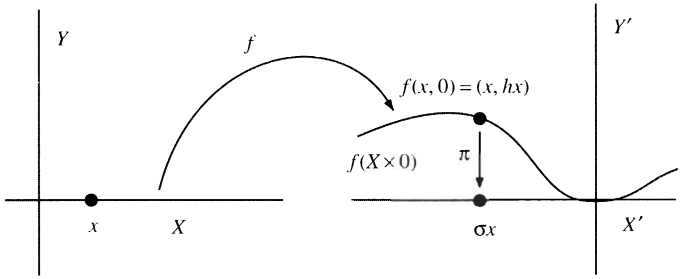


Figure 109 The image of $X \times 0$ is the graph of h.

For $(x', y') \in X' \times \mathbb{R}^{m-k}$, define

$$\psi(x', y') = (\sigma^{-1}(x'), y' - h(x')).$$

Since ψ is the composite of C^r diffeomorphisms,

$$(x', y') \mapsto (x', y' - h(x')) \mapsto (\sigma^{-1}(x'), y' - h(x')),$$

it too is a C^r diffeomorphism. (Alternately, you could compute the derivative of ψ at the origin and apply the Inverse Function Theorem.) We observe that $g = \psi \circ f \approx_r f$ satisfies

$$\begin{aligned} g(x, 0) &= \psi \circ (f_X(x, 0),\ f_Y(x, 0)) \\ &= (\sigma^{-1} \circ f_X(x, 0),\ f_Y(x, 0) - h(f_X(x, 0)) = (x, 0). \end{aligned}$$

Thus, it is no loss of generality to assume in the first place that $f(x, 0) = (x, 0)$. We do so.

Step 4. Finally, we find a local diffeomorphism φ in the neighborhood of 0 in $\mathbb{R}^n$ so that $f \circ \varphi$ is the projection map $P(x, y) = (x, 0)$.

The equation

$$f_X(\xi, y) - x = 0$$

defines $\xi = \xi(x, y)$ implicitly in a neighborhood of the origin; it is a C^r map from $\mathbb{R}^n$ into $\mathbb{R}^k$ and has $\xi(0, 0) = 0$. For, at the origin, the derivative

of $f_X(\xi, y) - x$ with respect to ξ is the invertible matrix $I_{k\times k}$. We claim that

$$\varphi(x, y) = (\xi(x, y),\ y)$$

is a local diffeomorphism of $\mathbb{R}^n$ and $G = f \circ \varphi$ is P.

The derivative of $\xi(x, y)$ with respect to x at the origin can be calculated from the chain rule (this was done in general for implicit functions) and it satisfies

$$0 = \frac{d\,F(\xi(x, y), x, y)}{dx} = \frac{\partial F}{\partial \xi}\frac{\partial \xi}{\partial x} + \frac{\partial F}{\partial x} = I_{k\times k}\frac{\partial \xi}{\partial x} - I_{k\times k}.$$

That is, at the origin $\partial\xi/\partial x$ is the identity matrix. Thus,

$$(D\varphi)_0 = \begin{bmatrix} I_{k\times k} & * \\ 0 & I_{(n-k)\times(n-k)} \end{bmatrix}$$

which is invertible no matter what $*$ is. Clearly $\varphi(0) = 0$. By the Inverse Function Theorem, φ is a local C^r diffeomorphism on a neighborhood of the origin and G is C^r equivalent to f. By Lemma 24, G has constant rank k.

We have

$$\begin{aligned} G(x, y) &= f \circ \varphi(x, y) = f(\xi(x, y),\ y) \\ &= (f_X(\xi, y),\ f_Y(\xi, y)) = (x,\ G_Y(x, y)). \end{aligned}$$

Therefore $G_X(x, y) = x$ and

$$DG = \begin{bmatrix} I_{k\times k} & 0 \\ * & \dfrac{\partial G_Y}{\partial y} \end{bmatrix}.$$

At last we use the constant rank hypothesis. (Until now, it has been enough that Df has rank $\geq k$.) The only way that a matrix of this form can have rank k is that

$$\frac{\partial G_Y}{\partial y} \equiv 0.$$

See Exercise 43. By Corollary 13 to the Mean Value Theorem, this implies that in a neighborhood of the origin, G_Y is independent of y. Thus

$$G_Y(x, y) = G_Y(x, 0) = f_Y(\xi(x, 0),\ 0),$$

which is 0 because $f_Y = 0$ on $\mathbb{R}^k \times 0$. The upshot is that $G \approx_r f$ and $G(x, y) = (x, 0)$; i.e., $G = P$. See also Exercise 31.

By Lemma 24, steps 1-4 concatenate to give a C^r equivalence between the original constant rank map f and the linear projection P. □

26 Corollary *If $f : U \to \mathbb{R}^m$ has rank k at p, then it is locally C^r equivalent to a map of the form $G(x, y) = (x, g(x, y))$ where $g : \mathbb{R}^n \to \mathbb{R}^{m-k}$ is C^r and $x \in \mathbb{R}^k$.*

Proof This was shown in the proof of the Rank Theorem before we used the assumption that f has constant rank k. □

27 Corollary *If $f : U \to \mathbb{R}$ is C^r and $(Df)_p$ has rank 1, then in a neighborhood of p the level sets $\{x \in U : f(x) = c\}$ form a stack of C^r nonlinear discs of dimension $n - 1$.*

Proof Near p the rank can not decrease, so f has constant rank 1 near p. The level sets of a projection $\mathbb{R}^n \to \mathbb{R}$ form a stack of planes and the level sets of f are the images of these planes under the equivalence diffeomorphism in the Rank Theorem. See Figure 110. □

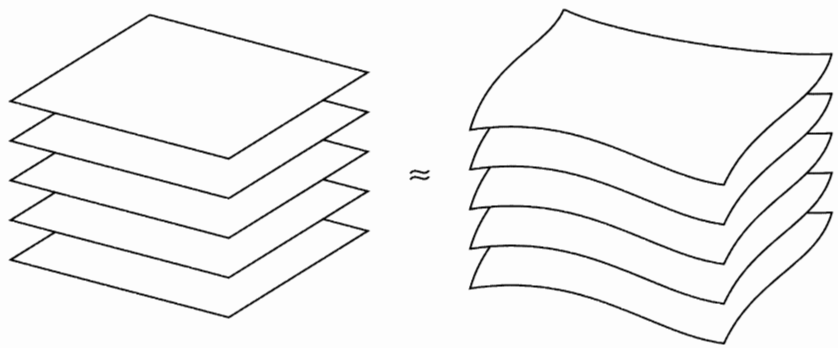


Figure 110 Near a rank-one point, the level sets of $f : U \to \mathbb{R}$ are diffeomorphic to a stack of planes.

28 Corollary *If $f : U \to \mathbb{R}^m$ has rank n at p then locally the image of U under f is a diffeomorphic copy of the n-dimensional disc.*

Proof Near p the rank can not decrease, so f has constant rank n near p. The Rank Theorem says that f is locally C^r equivalent to $x \mapsto (x, 0)$. (Since $k = n$, the y-coordinates are absent.) Thus, the local image of U is diffeomorphic to a neighborhood of 0 in $\mathbb{R}^n \times 0$ which is an n-dimensional disc. □

The geometric meaning of the diffeomorphisms ψ and φ is illustrated in the Figures 111 and 112.

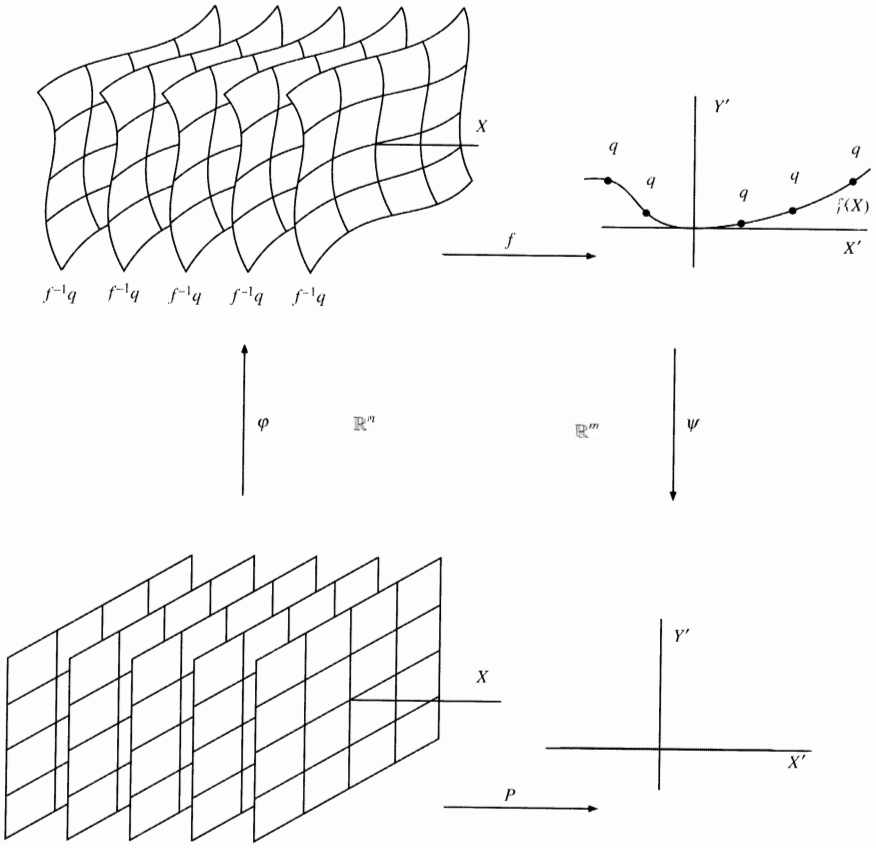


Figure 111 f has constant rank 1.

7* Lagrange Multipliers

In sophomore calculus you learn how to maximize a function $f(x, y, z)$ subject to a "constraint" or "side condition" $g(x, y, z) = \text{const.}$ by the Lagrange multiplier method. Namely, the maximum can occur only at a point p where the gradient of f is a scalar multiple of the gradient of g,

$$\operatorname{grad}_p f = \lambda \operatorname{grad}_p g.$$

The factor λ is the **Lagrange multiplier**. The goal of this section is a natural, mathematically complete explanation of the Lagrange multiplier method, which amounts to gazing at the right picture.

First, the natural hypotheses are:

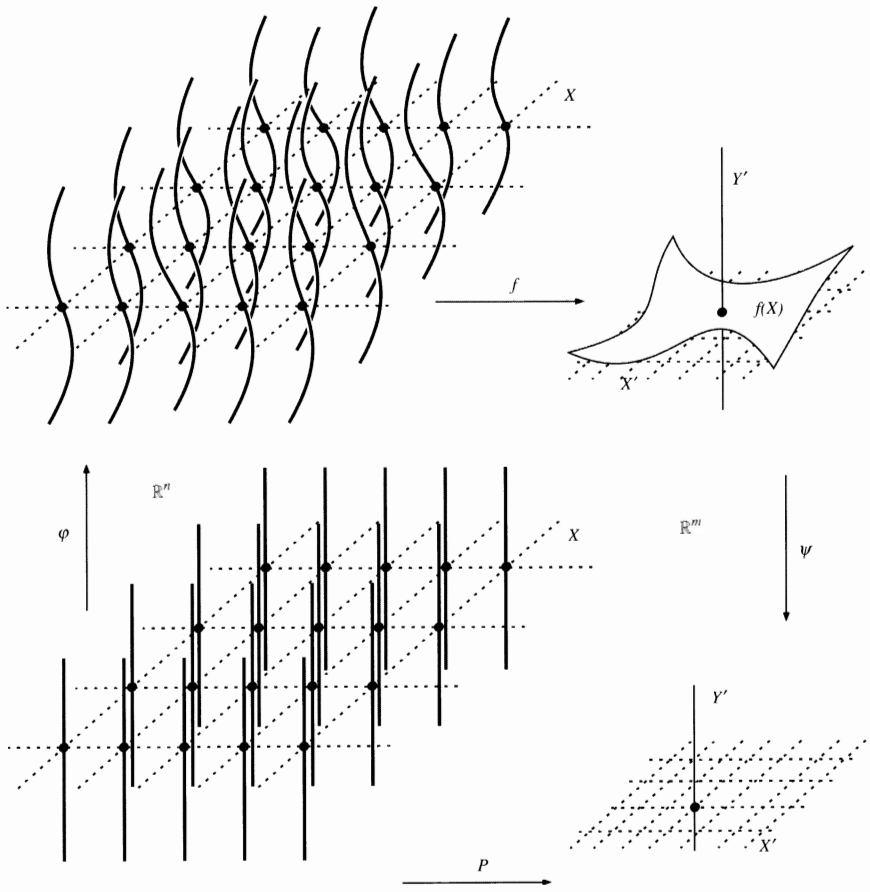


Figure 112 f has constant rank 2.

(a) f and g are C^1 real-valued functions defined on some region $U \subset \mathbb{R}^3$.

(b) For some constant c, the set $S = g^{\text{pre}}(c)$ is compact, nonempty, and $\operatorname{grad}_q g \neq 0$ for all $q \in S$.

The conclusion is

(c) The restriction of f to the set S, $f|_S$, has a maximum, say M, and if $p \in S$ has $f(p) = M$ then there is a λ such that $\operatorname{grad}_p f = \lambda \operatorname{grad}_p g$.

The method is utilized as follows. You are given† f and g, and you are asked to find a point $p \in S$ at which $f|_S$ is maximum. Compactness implies

† Sometimes you are merely given f and S. Then you must think up an appropriate g such that (b) is true.

that a maximum point exists, your job is to find it. You first locate all points $q \in S$ at which the gradients of f and g are linearly dependent; i.e., one gradient is a scalar multiple of the other. They are "candidates" for the maximum point. You then evaluate f at each candidate and the one with the largest f-value is the maximum. Done.

Of course you can find the minimum the same way. It too will be among the candidates, and it will have the smallest f-value. In fact the candidates are exactly the **critical points** of $f|_S$, the points $x \in S$ such that

$$\frac{fy - fx}{y - x} \to 0$$

as $y \in S$ tends to x.

Now we explain why the Lagrange multiplier method works. Recall that the **gradient** of a function $h(x, y, z)$ at $p \in U$ is the vector

$$\text{grad}_p\, h = \left(\frac{\partial h(p)}{\partial x}, \frac{\partial h(p)}{\partial y}, \frac{\partial h(p)}{\partial z}\right) \in \mathbb{R}^3.$$

Assume the hypotheses (a), (b), and that $f|_S$ attains its maximum value M at $p \in S$. We must prove (c) — the gradient of a f at p is a scalar multiple of the gradient of g at p. If $\text{grad}_p\, f = 0$ then $\text{grad}_p\, f = 0 \cdot \text{grad}_p\, g$, which verifies (c) degenerately. Thus, it is fair to assume that $\text{grad}_p\, f \neq 0$.

By the Rank Theorem, in the neighborhood of a point at which the gradient of f is nonzero, the f-level surfaces are like a stack of pancakes. (The pancakes are infinitely thin and may be somewhat curved. Alternately, you can picture the level surfaces as layers of an onion skin, or as a pile of transparency foils.)

To arrive at a contradiction, assume that $\text{grad}_p\, f$ is not a scalar multiple of $\text{grad}_p\, g$. The angle between the gradients is nonzero. Gaze at the f-level surfaces $f = M \pm \epsilon$ for ϵ small. The way these f-level surfaces meet the g-level surface S is shown in Figure 113.

The surface S is a knife blade that slices through the f-pancakes. The knife blade is perpendicular to grad g, while the pancakes are perpendicular to grad f. There is a positive angle between these gradient vectors, so the knife is not tangent to the pancakes. Rather, S slices transversely through each f-level surface near p, and $S \cap \{f = M + \epsilon\}$ is a curve that passes near p. The value of f on this curve is $M + \epsilon$, which contradicts the assumption that $f|_S$ attains a maximum at p. Therefore $\text{grad}_p\, f$ is, after all, a scalar multiple of $\text{grad}_p\, g$ and the proof of (c) is complete.

There is a higher-dimensional version of the Lagrange multiplier method. A C^1 function $f : U \to \mathbb{R}$ is defined on an open set $U \subset \mathbb{R}^n$, and it is constrained to a compact "surface" $S \subset U$ defined by k simultaneous

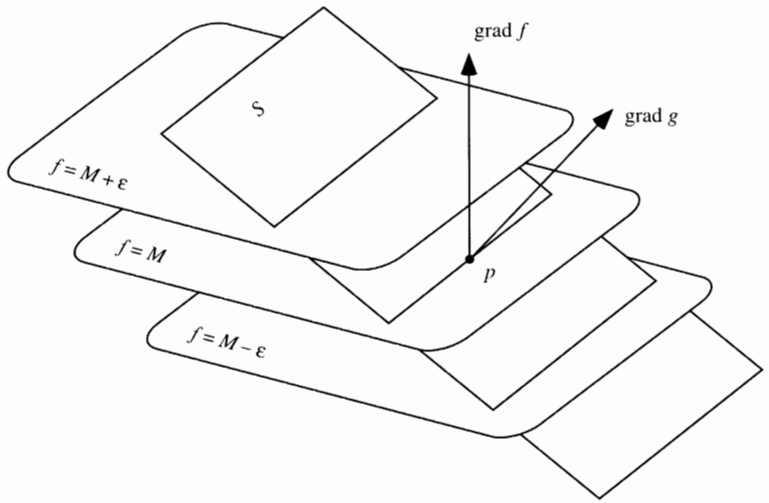


Figure 113 S cuts through all the f-level surfaces near p.

equations

$$g_1(x_1, \ldots, x_n) = c_1$$
$$\ldots$$
$$g_k(x_1, \ldots, x_n) = c_k.$$

We assume the functions g_i are C^1 and their gradients are linearly independent. The higher-dimensional Lagrange multiplier method asserts that if $f|_S$ achieves a maximum at p, then $\text{grad}_p\, f$ is a linear combination of $\text{grad}_p\, g_1, \ldots, \text{grad}_p\, g_k$. In contrast to Protter and Morrey's presentation on pages 369-372 of their book, *A First Course in Real Analysis*, the proof is utterly simple: it amounts to examining the situation in the right coordinate system at p.

It is no loss of generality to assume that p is the origin in $\mathbb{R}^n$ and that $c_1, \ldots, c_k, f(p)$ are zero. Also, we can assume that $\text{grad}_p\, f \neq 0$, since otherwise it is already a trivial linear combination of the gradients of the g_i. Then choose vectors $w_{k+2}, \ldots, w_n$ so that

$$\text{grad}_0\, g_1, \ldots, \text{grad}_0\, g_k, \quad \text{grad}_0\, f, \quad w_{k+2}, \ldots, w_n$$

is a basis of $\mathbb{R}^n$. For $k + 2 \leq i \leq n$ define

$$h_i(x) = \langle w_i, x \rangle.$$

The map $x \mapsto F(x) = (g_1(x), \ldots, g_k(x), \quad f(x), \quad h_{k+2}(x), \ldots, h_n(x))$ is a local diffeomorphism of $\mathbb{R}^n$ to itself, since the derivative of F at the origin is the $n \times n$ matrix of linearly independent column vectors,

$$(DF)_0 = \Big[\operatorname{grad}_0 g_1 \ldots \operatorname{grad}_0 g_k \quad \operatorname{grad}_0 f \quad w_{k+2} \ldots w_n \Big].$$

Think of the functions $y_i = F_i(x)$ as new coordinates on a neighborhood of the origin in $\mathbb{R}^n$. With respect to these coordinates, the surface S is the coordinate plane $0 \times \mathbb{R}^{n-k}$ on which the coordinates $y_1, \ldots, y_k$ are zero, and f is the $(k+1)^{\text{st}}$ coordinate function y_{k+1}. This coordinate function obviously does not attain a maximum on the coordinate plane $0 \times \mathbb{R}^{n-k}$, so $f|_S$ attains no maximum at p.

8 Multiple Integrals

In this section we generalize to n variables the one-variable Riemann integration theory appearing in Chapter 3. For simplicity, we assume throughout that the function f we integrate is real-valued, as contrasted to vector-valued, and at first we assume that f is a function of only two variables.

Consider a rectangle $R = [a, b] \times [c, d]$ in $\mathbb{R}^2$. Partitions P and Q of $[a, b]$ and $[c, d]$,

$$P : a = x_0 < x_1 < \cdots < x_m = b \qquad Q : c = y_0 < y_1 < \cdots < y_n = d,$$

give rise to a **grid** $G = P \times Q$ of rectangles

$$R_{ij} = I_i \times J_j$$

where $I_i = [x_{i-1}, x_i]$ and $J_j = [y_{j-1}, y_j]$. Let $\Delta x_i = x_i - x_{i-1}$, $\Delta y_j = y_j - y_{j-1}$, and denote the area of R_{ij} as

$$|R_{ij}| = \Delta x_i \, \Delta y_j.$$

Let S be a choice of sample points $(s_{ij}, t_{ij}) \in R_{ij}$. See Figure 114.

Given $f : R \to \mathbb{R}$, the corresponding **Riemann sum** is

$$R(f, G, S) = \sum_{i=1}^{m} \sum_{j=1}^{n} f(s_{ij}, t_{ij}) \, |R_{ij}| \, .$$

If there is a number to which the Riemann sums converge as the mesh of the grid (the diameter of the largest rectangle) tends to zero, then f is **Riemann integrable** and that number is the Riemann integral

$$\int_R f = \lim_{\text{mesh } G \to 0} R(f, G, S).$$

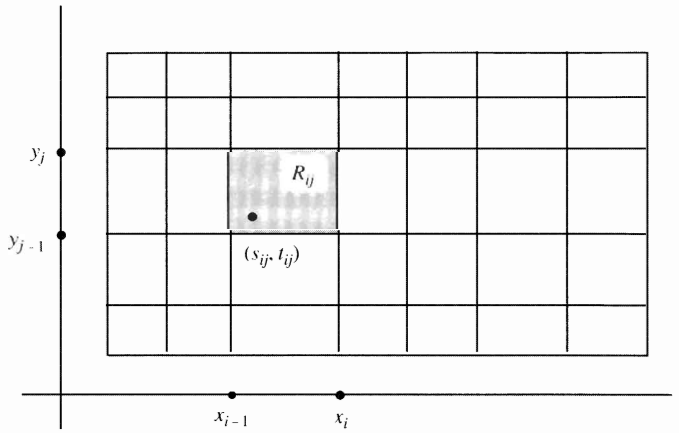


Figure 114 A grid and a sample point.

The lower and upper sums of a bounded function f with respect to the grid G are

$$L(f, G) = \sum m_{ij} \, |R_{ij}| \qquad U(f, G) = \sum M_{ij} \, |R_{ij}|$$

where m_{ij} and M_{ij} are the infimum and supremum of $f(s, t)$ as (s, t) varies over R_{ij}. The lower integral is the supremum of the lower sums and the upper integral is the infimum of the upper sums.

The proofs of the following facts are conceptually identical to the one-dimensional versions explained in Chapter 3:

(a) If f is Riemann integrable then it is bounded.

(b) The set of Riemann integrable functions $R \to \mathbb{R}$ is a vector space $\mathcal{R} = \mathcal{R}(R)$ and integration is a linear map $\mathcal{R} \to \mathbb{R}$.

(c) The constant function $f = k$ is integrable and its integral is $k \, |R|$.

(d) If $f, g \in \mathcal{R}$ and $f \leq g$ then

$$\int_R f \leq \int_R g.$$

(e) Every lower sum is less than or equal to every upper sum, and consequently the lower integral is no greater than the upper integral,

$$\underline{\int_R} f \leq \overline{\int_R} f.$$

(f) For a bounded function, Riemann integrability is equivalent to the equality of the lower and upper integrals, and integrability implies equality of the lower, upper, and Riemann integrals.

The Riemann-Lebesgue Theorem is another result that generalizes naturally to multiple integrals. It states that a bounded function is Riemann integrable if and only if its discontinuities form a zero set.

First of all, $Z \subset \mathbb{R}^2$ is a **zero set** if for each $\epsilon > 0$ there is a countable covering of Z by open rectangles S_k whose total area is $< \epsilon$,

$$\sum_k |S_k| < \epsilon.$$

By the $\epsilon/2^k$ construction, a countable union of zero sets is a zero set.

As in dimension one, we express the discontinuity set of our function $f : R \to \mathbb{R}$ as the union

$$D = \bigcup_{\kappa>0} D_\kappa,$$

where D_κ is the set points $z \in R$ at which the oscillation is $\geq \kappa$. That is,

$$\operatorname{osc}_z f = \lim_{r\to 0} \operatorname{diam}(f(R_r(z))) \geq \kappa$$

where $R_r(z)$ is the r-neighborhood of z in R. The set D_κ is compact.

Assume that $f : R \to \mathbb{R}$ is Riemann integrable. It is bounded and its upper and lower integrals are equal. Fix $\kappa > 0$. Given $\epsilon > 0$, there exists $\delta > 0$ such that if G is a grid with mesh $< \delta$ then

$$U(f, G) - L(f, G) < \epsilon.$$

Fix such a grid G. Each R_{ij} in the grid that contains in its interior a point of D_κ has $M_{ij} - m_{ij} \geq \kappa$, where m_{ij} and M_{ij} are the infimum and supremum of f on R_{ij}. The other points of D_κ lie in the zero set of gridlines $x_i \times [c, d]$ and $[a, b] \times y_j$. Since $U - L < \epsilon$, the total area of these rectangles with oscillation $\geq \kappa$ does not exceed ϵ/κ. Since κ is fixed and ϵ is arbitrary, D_κ is a zero set. Taking $\kappa = 1/2, 1/3, \ldots$ shows that the discontinuity set $D = \bigcup D_\kappa$ is a zero set.

Conversely, assume that f is bounded and D is a zero set. Fix any $\kappa > 0$. Each $z \in R \setminus D_\kappa$ has a neighborhood $W = W_z$ such that

$$\sup\{f(w) : w \in W\} - \inf\{f(w) : w \in W\} < \kappa.$$

Since D_κ is a zero set, it can be covered by countably many open rectangles S_k of small total area, say

$$\sum |S_k| < \sigma.$$

Let $\mathcal{V}$ be the covering of R by the neighborhoods W with small oscillation, and the rectangles S_k. Since R is compact, $\mathcal{V}$ has a positive Lebesgue number λ. Take a grid with mesh $< \lambda$. This breaks the sum

$$U - L = \sum (M_{ij} - m_{ij}) \, |R_{ij}|$$

into two parts: the sum of those terms for which R_{ij} is contained in a neighborhood W with small oscillation, plus a sum of terms for which R_{ij} is contained in one of the rectangles S_k. The latter sum is $< 2M\sigma$, while the former is $< \kappa \, |R|$. Thus, when κ and σ are small, $U - L$ is small, which implies Riemann integrability. To summarize,

The Riemann-Lebesgue Theorem remains valid for functions of several variables.

Now we come to the first place that multiple integration has something new to say. Suppose that $f : R \to \mathbb{R}$ is bounded and define

$$\underline{F}(y) = \underline{\int_a^b} f(x, y)\, dx \qquad \overline{F}(y) = \overline{\int_a^b} f(x, y)\, dx.$$

For each fixed $y \in [c, d]$, these are the lower and upper integrals of the single variable function $f_y : [a, b] \to \mathbb{R}$ defined by $f_y(x) = f(x, y)$. They are the integrals of $f(x, y)$ on the slice $y = \text{const}$. See Figure 115.

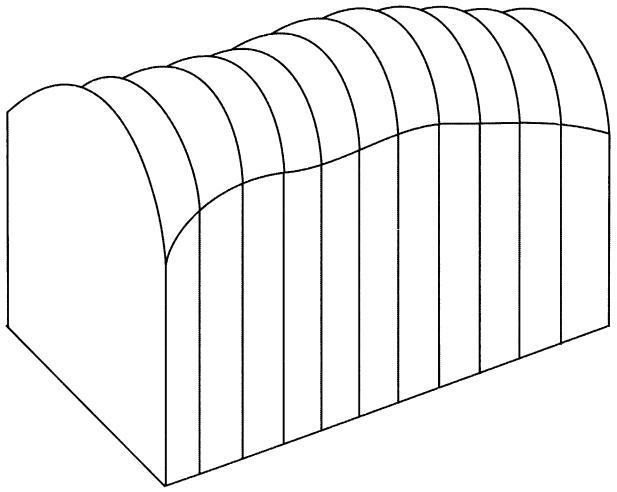


Figure 115 Fubini's Theorem is like sliced bread.

29 Fubini's Theorem *If f is Riemann integrable then so are $\underline{F}$ and $\overline{F}$. Moreover,*

$$\int_R f = \int_c^d \underline{F}\,dy = \int_c^d \overline{F}\,dy.$$

Since $\underline{F} \le \overline{F}$ and the integral of their difference is zero, it follows from the one-dimensional Riemann-Lebesgue Theorem that there exists a linear zero set $Y \subset [c, d]$, such that if $y \notin Y$ then $\underline{F}(y) = \overline{F}(y)$. That is, the integral of $f(x, y)$ with respect to x exists for almost all y, and we get the more common way to write the Fubini formula

$$\iint_R f\,dxdy = \int_c^d \Big[\int_a^b f(x, y)\,dx\Big]dy.$$

There is, however, an ambiguity in this formula. What is the value of the integrand $\int_a^b f(x, y)\,dx$ when $y \in Y$? For such a y, $\underline{F}(y) < \overline{F}(y)$ and the integral of $f(x, y)$ with respect to x does not exist. The answer is that we can choose any value between $\underline{F}(y)$ and $\overline{F}(y)$. The integral with respect to y will be unaffected. See also Exercise 47.

Proof We claim that if P and Q are partitions of $[a, b]$ and $[c, d]$ then

$$L(f, G) \le L(\underline{F}, Q) \tag{9}$$

where G is the grid $P \times Q$. Fix any partition interval $J_j \subset [c, d]$. If $y \in J_j$ then

$$m_{ij} = \inf\{f(s, t) : (s, t) \in R_{ij}\} \le \inf\{f(s, y) : s \in I_i\} = m_i(f_y).$$

Thus

$$\sum_{i=1}^m m_{ij}\Delta x_i \le \sum_{i=1}^m m_i(f_y)\Delta x_i = L(f_y, P) \le \underline{F}(y),$$

and it follows that

$$\sum_{i=1}^m m_{ij}\Delta x_i \le m_j(\underline{F}).$$

Therefore

$$\sum_{j=1}^n \sum_{i=1}^m m_{ij}\Delta x_i \Delta y_j \le \sum_{j=1}^n m_j(\underline{F})\Delta y_j = L(\underline{F}, Q)$$

which gives (9). Analogously, $U(\overline{F}, Q) \le U(f, G)$. Thus

$$L(f, G) \le L(\underline{F}, Q) \le U(\underline{F}, Q) \le U(\overline{F}, Q) \le U(f, G).$$

Since f is integrable, the outer terms of this inequality differ by arbitrarily little when the mesh of G is small. Taking infima and suprema over all grids $G = P \times Q$ gives

$$\int_R f = \sup L(f, G) \le \sup L(\underline{F}, Q) \le \inf U(\underline{F}, Q)$$
$$\le \inf U(f, G) = \int_R f.$$

The resulting equality of these five quantities implies that $\underline{F}$ is integrable and its integral on $[c, d]$ equals that of f on R. The case of the upper integral is handled in the same way. □

30 Corollary *If f is Riemann integrable, then the order of integration — first x then y, or vice versa — is irrelevant to the value of the iterated integral,*

$$\int_c^d \Big[\int_a^b f(x, y)\,dx\Big]dy = \int_a^b \Big[\int_c^d f(x, y)\,dy\Big]dx.$$

Proof Both iterated integrals equal the integral of f over R. □

A geometric consequence of Fubini's Theorem concerns the calculation of the area of plane regions by a slice method. Corresponding slice methods are valid in 3-space and in higher dimensions.

31 Cavalieri's Principle *The area of a region $S \subset R$ is the integral with respect to x of the length of its vertical slices,*

$$\text{area}(S) = \int_a^b \text{length}(S_x)\,dx,$$

provided that the boundary of S is a zero set.

See Appendix B for a delightful discussion of the historical origin of Cavalieri's Principle. Deriving Cavalieri's Principle is mainly a matter of definition. For we define the length of a subset of $\mathbb{R}$ and the area of a subset of $\mathbb{R}^2$ to be the integrals of their characteristic functions. The requirement that ∂S is a zero set is made so that χ_S is Riemann integrable. It is met if S has a smooth, or piecewise smooth, boundary. See Chapter 6 for a more geometric definition of length and area in terms of outer measure.

The second new aspect of multiple integration concerns the change of variables formula. We will suppose that $\varphi : U \to W$ is a C^1 diffeomorphism between open subsets of $\mathbb{R}^2$, that $R \subset U$, and that a Riemann

integrable function $f : W \to \mathbb{R}$ is given. The **Jacobian** of φ at $z \in U$ is the determinant of the derivative,

$$\mathrm{Jac}_z\, \varphi = \det(D\varphi)_z.$$

32 Change of Variables Formula *Under the preceding assumptions*

$$\int_R f \circ \varphi \; |\mathrm{Jac}\, \varphi| \; = \int_{\varphi(R)} f.$$

See Figure 116.

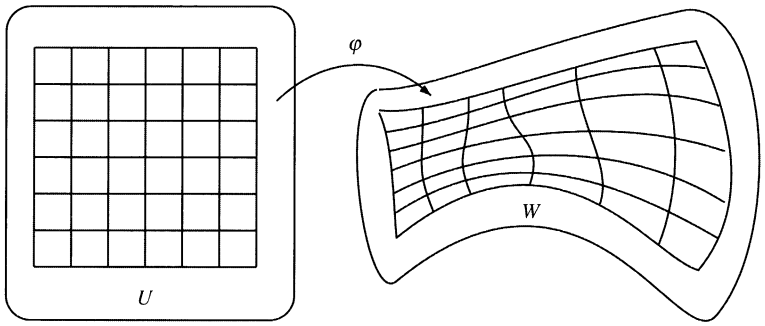


Figure 116 φ is a change of variables.

If S is a bounded subset of $\mathbb{R}^2$, its **area** (or **Jordan content**) is by definition the integral of its characteristic function χ_S, if the integral exists; when the integral does exist we say that S is **Riemann measurable**. See also Appendix B of Chapter 6. According to the Riemann-Lebesgue Theorem, S is Riemann measurable if and only if its boundary is a zero set. For χ_S is discontinuous at z if and only if z is a boundary point of S. See Exercise 44. The characteristic function of a rectangle R is Riemann integrable, its integral is $|R|$, so we are justified in using the same notation for area of a general set S, namely

$$|S| = \mathrm{area}(S) = \int \chi_S.$$

33 Proposition *If $T : \mathbb{R}^2 \to \mathbb{R}^2$ is an isomorphism then for every Riemann measurable set $S \subset \mathbb{R}^2$, $T(S)$ is Riemann measurable and*

$$|T(S)| = |\det T|\, |S|\,.$$

Proposition 33 is a version of the Change of Variables Formula in which $\varphi = T$, $R = S$, and $f = 1$. It remains true for n-dimensional volume and leads to a *definition* of the determinant of a linear transformation as a "volume multiplier."

Proof As is shown in linear algebra, the matrix A that represents T is a product of elementary matrices

$$A = E_1 \cdots E_k.$$

Each elementary 2×2 matrix is one of the following types:

$$\begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & \sigma \\ 0 & 1 \end{bmatrix}$$

where $\lambda > 0$. The first three matrices represent isomorphisms whose effect on I^2 is obvious: I^2 is converted to the rectangles $\lambda I \times I$, $I \times \lambda I$, I^2. In each case, the area agrees with the magnitude of the determinant. The fourth isomorphism converts I^2 to the parallelogram

$$\Pi = \{(x, y) \in \mathbb{R}^2 : \sigma y \le x \le 1 + \sigma y \text{ and } 0 \le y \le 1\}.$$

Π is Riemann measurable since its boundary is a zero set. By Fubini's Theorem, we get

$$|\Pi| = \int \chi_\Pi = \int_0^1 \Big[\int_{x=\sigma y}^{x=1+\sigma y} 1\, dx \Big] dy = 1 = \det E.$$

Exactly the same thinking shows that for any rectangle R, not merely the unit square,

$$|E(R)| = |\det E|\, |R|\,. \tag{10}$$

We claim that (10) implies that for any Riemann measurable set S, $E(S)$ is Riemann measurable and

$$|E(S)| = |\det E|\, |S|\,. \tag{11}$$

Let $\epsilon > 0$ be given. Choose a grid G on $R \supset S$ with mesh so small that the rectangles R of G satisfy

$$|S| - \epsilon \; \le \sum_{R \subset S} |R| \; \le \sum_{R \cap S \neq \emptyset} |R| \; \le \; |S| + \epsilon. \tag{12}$$

The interiors of the inner rectangles — those with $R \subset S$ — are disjoint, and therefore for each $z \in \mathbb{R}^2$,

$$\sum_{R \subset S} \chi_{\mathrm{int}\, R}(z) \le \chi_S(z).$$

The same is true after we apply E, namely

$$\sum_{R\subset S} \chi_{\text{int}(E(R))}(z) \le \chi_{E(S)}(z).$$

Linearity and monotonicity of the integral, and Riemann measurability of the sets $E(R)$ imply that

$$(13)\qquad \sum_{R\subset S} |E(R)| = \sum_{R\subset S} \int \chi_{\text{int}(E(R))} = \sum_{R\subset S} \underline{\int} \chi_{\text{int}(E(R))} \le \underline{\int} \chi_{E(S)}.$$

Similarly,

$$\chi_{E(S)}(z) \le \sum_{R\cap S\neq\emptyset} \chi_{E(R)}(z),$$

which implies that

$$(14)\qquad \overline{\int} \chi_{E(S)} \le \sum_{R\cap S\neq\emptyset} \overline{\int} \chi_{E(R)} = \sum_{R\cap S\neq\emptyset} \int \chi_{E(R)} = \sum_{R\cap S\neq\emptyset} |E(R)|.$$

By (10) and (12), (13) and (14) become

$$\begin{aligned} |\det E|\,(|S|-\epsilon) &\le |\det E| \sum_{R\subset S} |R| \\ &\le \underline{\int} \chi_{E(S)} \le \overline{\int} \chi_{E(S)} \le |\det E| \sum_{R\cap S\neq\emptyset} |R| \\ &\le |\det E|\,(|S|+\epsilon). \end{aligned}$$

Since these upper and lower integrals do not depend on ϵ, and ϵ is arbitrarily small, they equal the common value $|\det E|\,|S|$, which completes the proof of (11).

The determinant of a matrix product is the product of the determinants. Since the matrix of T is the product of elementary matrices, $E_1 \cdots E_k$, (11) implies that if S is Riemann measurable then so is $T(S)$ and

$$\begin{aligned} |T(S)| &= |E_1 \cdots E_k(S)| \\ &= |\det E_1| \cdots |\det E_k|\,|S| = |\det T|\,|S|. \end{aligned}$$

□

We isolate two more facts in preparation for the proof of the Change of Variables Formula.

34 Lemma *Suppose that $\psi : U \to \mathbb{R}^2$ is C^1, $0 \in U$, $\psi(0) = 0$, and for all $u \in U$,*

$$\|(D\psi)_u - \text{Id}\| \le \epsilon.$$

If $U_r(0) \subset U$ then

$$\psi(U_r(0)) \subset U_{(1+\epsilon)r}(0).$$

Proof By $U_r(p)$ we denote the r-neighborhood of p in U. The C^1 Mean Value Theorem gives

$$\begin{aligned} \psi(u) = \psi(u) - \psi(0) &= \int_0^1 (D\psi)_{tu}\,dt\,(u) \\ &= \int_0^1 \Big((D\psi)_{tu} - \text{Id}\Big)dt\,(u) + u. \end{aligned}$$

If $|u| \le r$ this implies that $|\psi(u)| \le (1+\epsilon)r$; i.e., $\psi(U_r(0)) \subset U_{(1+\epsilon)r}(0)$. □

Lemma 34 is valid for any choice of norm on $\mathbb{R}^2$, in particular for the maximum coordinate norm. In that case, the inclusion refers to squares: the square of radius r is carried by ψ inside the square of radius $(1+\epsilon)r$.

35 Lemma *The Lipschitz image of a zero set is a zero set.*

Proof Suppose that Z is a zero set and $h : Z \to \mathbb{R}^2$ satisfies a Lipschitz condition

$$\left|h(z) - h(z')\right| \le L\left|z - z'\right|.$$

Given $\epsilon > 0$, there is a countable covering of Z by squares S_k such that

$$\sum_k |S_k| < \epsilon.$$

See Exercise 45. Each set $S_k \cap Z$ has diameter $\le$ diam S_k, and therefore $h(Z \cap S_k)$ has diameter $\le L$ diam S_k. As such, it is contained in a square S_k' of edge length L diam S_k. The squares S_k' cover $h(Z)$ and

$$\sum_k |S_k'| \le L^2 \sum_k (\text{diam } S_k)^2 = 2L^2 \sum_k |S_k| \le 2L^2\epsilon.$$

Therefore $h(Z)$ is a zero set. □

Proof of the Change of Variables Formula Recall that $\varphi : U \to W$ is a C^1 diffeomorphism, $f : W \to \mathbb{R}$ is Riemann integrable, R is a rectangle in U, and it is asserted that

$$\int_R f \circ \varphi \cdot |\operatorname{Jac} \varphi| = \int_{\varphi(R)} f. \tag{15}$$

Let D' be the set of discontinuity points of f. It is a zero set. Then

$$D = \varphi^{-1}(D')$$

is the set of discontinuity points of $f \circ \varphi$. The C^1 Mean Value Theorem implies that φ^{-1} is Lipschitz, Lemma 35 implies that D is a zero set, and the Riemann-Lebesgue Theorem implies that $f \circ \varphi$ is Riemann integrable. Since $|\operatorname{Jac} \varphi|$ is continuous, it is Riemann integrable, and so is the product $f \circ \varphi \cdot |\operatorname{Jac} \varphi|$. In short, the l.h.s. of (15) makes sense.

Since φ is a diffeomorphism, it is a homeomorphism and it carries the boundary of R to the boundary of $\varphi(R)$. The former boundary is a zero set and, by Lemma 35, so is the latter. Thus $\chi_{\varphi(R)}$ is Riemann integrable. Choose a rectangle R' that contains $\varphi(R)$. Then the r.h.s. of (15) becomes

$$\int_{\varphi(R)} f = \int_{R'} f \cdot \chi_{\varphi(R)},$$

which also makes sense. It remains to show that the two sides of (15) not only make sense, but are equal.

Equip $\mathbb{R}^2$ with the maximum coordinate norm and equip $\mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$ with the associated operator norm

$$\|T\| = \max\{|T(v)|_{\max} : |v|_{\max} \leq 1\}.$$

Let $\epsilon > 0$ be given. Take any grid G that partitions R into squares R_{ij} of radius r. (The smallness of r will be specified below.) Let z_{ij} be the center point of R_{ij} and call

$$A_{ij} = (D\varphi)_{z_{ij}} \qquad \varphi(z_{ij}) = w_{ij} \qquad \varphi(R_{ij}) = W_{ij}.$$

The Taylor approximation to φ on R_{ij} is

$$\phi_{ij}(z) = w_{ij} + A_{ij}(z - z_{ij}).$$

The composite $\psi = \phi_{ij}^{-1} \circ \varphi$ sends z_{ij} to itself and its derivative at z_{ij} is the identity transformation. Uniform continuity of $(D\varphi)_z$ on R implies that

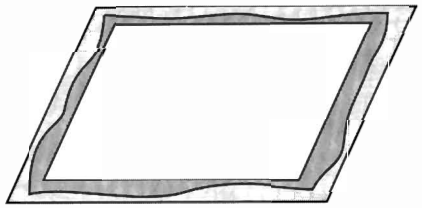


Figure 117 How we magnify the picture and sandwich a nonlinear parallelogram between two linear ones.

if r is small enough then for all $z \in R_{ij}$ and for all ij, $\|(D\psi)_z - \mathrm{Id}\| < \epsilon$. By Lemma 34,

$$\phi_{ij}^{-1} \circ \varphi(R_{ij}) \subset (1+\epsilon) R_{ij} \tag{16}$$

where $(1+\epsilon)R_{ij}$ refers to the $(1+\epsilon)$-dilation of R_{ij} centered at z_{ij}. Similarly, Lemma 34 applies to the composite $\varphi^{-1} \circ \phi_{ij}$ and, taking the radius $r/(1+\epsilon)$ instead of r, we get

$$\varphi^{-1} \circ \phi_{ij}((1+\epsilon)^{-1} R_{ij}) \subset R_{ij}. \tag{17}$$

See Figure 117. Then (16) and (17) imply

$$\phi_{ij}((1+\epsilon)^{-1} R_{ij}) \subset \varphi(R_{ij}) = W_{ij} \subset \phi_{ij}((1+\epsilon) R_{ij}).$$

By Proposition 33, this gives the area estimate

$$\frac{J_{ij} \left|R_{ij}\right|}{(1+\epsilon)^2} \leq \left|W_{ij}\right| \leq (1+\epsilon)^2 J_{ij} \left|R_{ij}\right|,$$

where $J_{ij} = \left|\operatorname{Jac}_{z_{ij}} \varphi\right|$. Equivalently,

$$\frac{1}{(1+\epsilon)^2} \leq \frac{\left|W_{ij}\right|}{J_{ij} \left|R_{ij}\right|} \leq (1+\epsilon)^2. \tag{18}$$

An estimate of the form

$$\frac{1}{(1+\epsilon)^2} \leq \frac{a}{b} \leq (1+\epsilon)^2,$$

with $0 \leq \epsilon \leq 1$ and $a, b > 0$, implies that

$$|a - b| \leq 16\epsilon b,$$

as you are left to check in Exercise 40. Thus (18) implies

$$\left|\left|W_{ij}\right| - J_{ij} \left|R_{ij}\right|\right| \leq 16\epsilon J \left|R_{ij}\right| \tag{19}$$

where $J = \sup\{|\mathrm{Jac}_z\,\varphi| : z \in R\}$.

Let m_{ij} and M_{ij} be the infimum and supremum of $f \circ \varphi$ on R_{ij}. Then, for all $w \in \varphi(R)$,

$$\sum m_{ij} \chi_{\operatorname{int} W_{ij}}(w) \le f(w) \le \sum M_{ij} \chi_{W_{ij}}(w),$$

which integrates to

$$\sum m_{ij} \left|W_{ij}\right| \le \int_{\varphi(R)} f \le \sum M_{ij} \left|W_{ij}\right|.$$

According to (19), replacing $\left|W_{ij}\right|$ by $J_{ij} \left|R_{ij}\right|$ causes an error of no more than $16\epsilon J \left|R_{ij}\right|$. Thus

$$\sum m_{ij} J_{ij} \left|R_{ij}\right| - 16\epsilon M J\,|R| \le \int_{\varphi(R)} f \le \sum M_{ij} J_{ij} \left|R_{ij}\right| + 16\epsilon M J\,|R|,$$

where $M = \sup|f|$. These are lower and upper sums for the integrable function $f \circ \varphi \cdot |\mathrm{Jac}\,\varphi|$. Thus

$$\int_R f\circ\varphi\cdot|\mathrm{Jac}\,\varphi| - 16\epsilon M J\,|R| \le \int_{\varphi(R)} f \le \int_R f\circ\varphi\cdot|\mathrm{Jac}\,\varphi| + 16\epsilon M J\,|R|.$$

Since ϵ is arbitrarily small, the proof is complete. □

Finally, here is a sketch of the n-dimensional theory. Instead of a two-dimensional rectangle we have a box

$$R = [a_1, b_1] \times \cdots \times [a_n, b_n].$$

Riemann sums of a function $f : R \to \mathbb{R}$ are defined as before: take a grid G of small boxes R_ℓ in R, take a sample point s_ℓ in each, and set

$$R(f, G, S) = \sum f(s_\ell)\,|R_\ell|$$

where $|R_\ell|$ is the product of the edge lengths of the small box R_ℓ and S is the set of sample points. If the Riemann sums converge to a limit it is the integral. The general theory, including the Riemann-Lebesgue Theorem, is the same as in dimension two.

Fubini's Theorem is proved by induction on n, and has the same meaning: integration on a box can be done slice by slice, and the order in which the iterated integration is performed has no effect on the answer.

The Change of Variables Formula has the same statement, only now the Jacobian is the determinant of an $n \times n$ matrix. In place of area we have volume, the n-dimensional volume of a set $S \subset \mathbb{R}^n$ being the integral of its characteristic function. The volume-multiplier formula, Proposition 33, has essentially the same proof, but the elementary matrix notation is messier. (It helps to realize that the following types of elementary row operations suffice for row reduction: transposition of two adjacent rows, multiplication of the first row by λ, and addition of the second row to the first.) The proof of the Change of Variables Formula itself differs only in that 16 becomes 4^n.

9 Differential Forms

The Riemann integral notation

$$\sum_{i=1}^{n} f(t_i)\,\Delta x_i \quad \approx \quad \int_a^b f(x)\,dx$$

may lead one to imagine the integral as an "infinite sum of infinitely small quantities $f(x)dx$." Although this idea itself seems to lead nowhere, it points to a good question — how do you give an independent meaning to the symbol $f\,dx$? The answer: differential forms. Not only does the theory of differential forms supply coherent, independent meanings for $f\,dx$, dx, dy, df, $dx\,dy$, and even for d and x separately, it also unifies vector calculus results. A single result, the General Stokes Formula for differential forms

$$\int_M d\omega = \int_{\partial M} \omega,$$

encapsulates all integral theorems about divergence, gradient, and curl.

The presentation of differential forms in this section appears in the natural generality of n dimensions, and as a consequence it is unavoidably fraught with complicated index notation — armies of i's, j's, double subscripts, multi-indices, and so on. Your endurance may be tried.

First, consider a function $y = F(x)$. Normally, you think of F as the function, x as the input variable, and y as the output variable. But you can also take a *dual* approach and think of x as the function, F as the input variable, and y as the output variable. After all, why not? It's a kind of mathematical yin/yang.

Now consider a path integral the way it is defined in calculus,

$$\int_C f\,dx + g\,dy = \int_0^1 f(x(t), y(t))\frac{dx(t)}{dt}\,dt + \int_0^1 g(x(t), y(t))\frac{dy(t)}{dt}\,dt.$$

f and g are smooth real-valued functions of (x, y) and C is a smooth path parameterized by $(x(t), y(t))$ as t varies on $[0, 1]$. Normally, you think of the integral as a number that depends on the functions f and g. Taking the dual approach, you can think of it as a number that depends on the path C. This will be our point of view. It parallels that found in Rudin's *Principles of Mathematical Analysis*.

Definition A **differential 1-form** is a function that sends paths to real numbers and which can be expressed as a path integral in the previous notation. The **name** of this particular differential 1-form is $f\,dx + g\,dy$.

In a way, this definition begs the question. For it simply says that the standard calculus formula for path integrals should be read in a new way — as a function of the integration domain. Doing so, however, is illuminating, for it leads you to ask: just what property of C does the differential 1-form $f\,dx + g\,dy$ measure?

First take the case that $f(x, y) = 1$ and $g(x, y) = 0$. Then the path integral is

$$\int_C dx = \int_a^b \frac{dx(t)}{dt}\,dt = x(b) - x(a)$$

which is the **net x-variation** of the path C. This can be written in functional notation as

$$dx : C \mapsto x(b) - x(a).$$

It means that dx assigns to each path C its net x-variation. Similarly dy assigns to each path its net y-variation. The word "net" is important. Negative x-variation cancels positive x-variation, and negative y-variation cancels positive y-variation. In the world of forms, orientation matters.

What about $f\,dx$? The function f "weights" x-variation. If the path C passes through a region in which f is large, its x-variation is magnified accordingly, and the integral $\int_C f\,dx$ reflects the net f-weighted x-variation of C. In functional notation

$$f\,dx : C \mapsto \text{ net } f\text{-weighted } x\text{-variation of } C.$$

Similarly, $g\,dy$ assigns to a path its net g-weighted y-variation, and the 1-form $f\,dx + g\,dy$ assigns to C the sum of the two variations.

Terminology A **functional** on a space X is a function from X to $\mathbb{R}$.

Differential 1-forms are functionals on the space of paths. Some functionals on the space of paths are differential forms but others are not. For

instance, assigning to each path its arc-length is a functional that is not a form. For if C is a path parameterized by $(x(t), y(t))$, then $(x^*(t), y^*(t)) = (x(a+b-t), y(a+b-t))$ parameterizes C in the reverse direction. Arc-length is unaffected, but the value of any 1-form on the path changes sign. Hence, arc-length is not a 1-form. A more trivial example is the functional that assigns to each path the number 1. It too fails to have the right symmetry property under parameter reversal, and is not a form.

The definition of k-forms for $k \geq 2$ requires Jacobian determinants. To simplify notation we write $\partial F_I/\partial x_J = \partial(F_{i_1}, \ldots, F_{i_k})/\partial(x_{j_1}, \ldots, x_{j_k})$ where $I = (i_1, \ldots, i_k)$, $J = (j_1, \ldots, j_k)$ are k-tuples of integers, and $F : \mathbb{R}^n \to \mathbb{R}^m$ is smooth. Thus,

$$\frac{\partial F_I}{\partial x_J} = \det \begin{bmatrix} \dfrac{\partial F_{i_1}}{\partial x_{j_1}} & \cdots & \dfrac{\partial F_{i_1}}{\partial x_{j_k}} \\ \cdots & \cdots & \cdots \\ \dfrac{\partial F_{i_k}}{\partial x_{j_1}} & \cdots & \dfrac{\partial F_{i_k}}{\partial x_{j_k}} \end{bmatrix}$$

If $k = 1$, $I = (i)$, and $J = (j)$ then $\partial f_I/\partial x_J$ is just $\partial f_i/\partial x_j$.

Definition A **k-cell** in $\mathbb{R}^n$ is a smooth map $\varphi : I^k \to \mathbb{R}^n$ where I^k is the k-cube. If $I = (i_1, \ldots, i_k)$ is a k-tuple in $\{1, \ldots, n\}$ then dx_I is the functional that assigns to each k-cell φ its **x_I-area**

$$dx_I : \varphi \mapsto \int_{I^k} \frac{\partial \varphi_I}{\partial u}\,du$$

where this integral notation is shorthand for

$$\int_0^1 \cdots \int_0^1 \frac{\partial(\varphi_{i_1}, \ldots, \varphi_{i_k})}{\partial(u_1, \ldots, u_k)}\,du_1 \ldots du_k.$$

If f is a smooth function on $\mathbb{R}^n$ then $f\,dx_I$ is the functional

$$f\,dx_I : \varphi \mapsto \int_{I^k} f(\varphi(u))\frac{\partial \varphi_I}{\partial u}\,du.$$

The function f **weights** x_I-area. The functional dx_I is a **basic k-form** and $f\,dx_I$ is a **simple k-form**, while a sum of simple k-forms, is a (general) **k-form**

$$\omega = \sum_I f_I\,dx_I : \varphi \mapsto \sum_I (f_I\,dx_I)(\varphi).$$

The careful reader will detect some abuse of notation. Here I is used to index a collection of scalar coefficient functions $\{f_I\}$, whereas I is also used to reduce an m-vector $(F_1, \ldots, F_m)$ to a k-vector $F_I = (F_{i_1}, \ldots, F_{i_k})$. Besides this, I is the unit interval. Please persevere.

To underline the fact that a form is an integral, we write

$$\omega(\varphi) = \int_\varphi \omega.$$

Notation $C_k(\mathbb{R}^n)$ is the set of all k-cells in $\mathbb{R}^n$, $C^k(\mathbb{R}^n)$ is the set of all functionals on $C_k(\mathbb{R}^n)$, and $\Omega^k(\mathbb{R}^n)$ is the set of k-forms on $\mathbb{R}^n$.

Because a determinant changes sign under a row transposition, k-forms satisfy the **signed commutativity** property: if π permutes I to πI then

$$dx_{\pi I} = \operatorname{sgn}(\pi) dx_I.$$

In particular, $dx_{(1,2)} = -dx_{(2,1)}$. Because a determinant is zero if it has a repeated row, $dx_I = 0$ if I has a repeated entry. In particular $dx_{(1,1)} = 0$.

In terms of line integrals in the plane, paths are 1-cells, $x_{(1)} = x, x_{(2)} = y$, $dx_{(1)}$-area is net x-variation, and $dx_{(2)}$-area is net y-variation. Similarly, parameterized surface integrals (as discussed in sophomore calculus) correspond to integrals of 2-forms over 2-cells. The $x_{(1,2)}$-area of a surface is the net area of its projection to the xy-plane. The equation $dx_{(1,2)} = -dx_{(2,1)}$ signifies that xy-area is the negative of yx-area.

Form Naturality

It is a common error to confuse a cell, which a smooth mapping, with its image, which is point set — but the error is fairly harmless.

36 Theorem *Integrating a k-form over k-cells that differ by a reparameterization produces the same answer up to a factor of ± 1, and this factor of ± 1 is determined by whether the reparameterization preserves or reverses orientation.*

Proof If T is an orientation preserving diffeomorphism of I^k to itself and $\omega = f dx_I$ then the Jacobian $\partial T/\partial u$ is positive. The product determinant

formula and the change of variables formula for multiple integrals give

$$\begin{aligned}\int_{\varphi\circ T} \omega &= \int_{I^k} f(\varphi \circ T(u)) \frac{\partial(\varphi \circ T)_I}{\partial u}\, du \\ &= \int_{I^k} f(\varphi \circ T(u)) \Big(\frac{\partial \varphi_I}{\partial v}\Big)_{v=T(u)} \frac{\partial T}{\partial u}\, du \\ &= \int_{I^k} f(\varphi(v)) \frac{\partial \varphi_I}{\partial v}\, dv = \int_\varphi \omega.\end{aligned}$$

Taking sums shows that the equation $\int_{\varphi\circ T} \omega = \int_\varphi \omega$ continues to hold for all $\omega \in \Omega^k$. If T reverses orientation its Jacobian is negative. In the change of variables formula appears the absolute value of the Jacobian, which causes $\int_{\varphi\circ T} \omega$ to change sign. □

A particular case of the previous theorem concerns line integrals in the plane. The integral of a 1-form over a curve C does not depend on how C is parameterized. If we first parameterize C using a parameter $t \in [0, 1]$ and then reparameterize it by arc-length $s \in [0, L]$ where L is the length of C and the orientation of C remains the same, then integrals of 1-forms are unaffected,

$$\begin{aligned}\int_0^1 f(x(t), y(t)) \frac{dx(t)}{dt}\, dt &= \int_0^L f(x(s), y(s)) \frac{dx(s)}{ds}\, ds \\ \int_0^1 g(x(t), y(t)) \frac{dy(t)}{dt}\, dt &= \int_0^L g(x(s), y(s)) \frac{dy(s)}{ds}\, ds.\end{aligned}$$

Form Names

A k-tuple $I = (i_1, \ldots, i_k)$ **ascends** if $i_1 < \cdots < i_k$.

37 Proposition *Each k-form ω has a unique expression as a sum of simple k-forms with ascending k-tuple indices,*

$$\omega = \sum f_A dx_A.$$

Moreover, the coefficient $f_A(x)$ in this ***ascending presentation*** *(or "name") of ω is determined by the value of ω on small k-cells at x.*

Proof Using the signed commutativity property of forms, we regroup and combine a sum of simple forms into terms in which the indices ascend. This gives the existence of an ascending presentation $\omega = \sum f_A dx_A$.

Fix an ascending k-tuple A and fix a point $p \in \mathbb{R}^n$. For $r > 0$ consider the **inclusion cell**,

$$\iota = \iota_{r,p} : u \mapsto p + rL(u)$$

where L is the linear inclusion map that sends $\mathbb{R}^k$ to the x_A-plane. ι sends I^k to a cube in the x_A-plane at p. As $r \to 0$, the cube shrinks to p. If I ascends, the Jacobians of ι are

$$\frac{\partial \iota_I}{\partial u} = \begin{cases} r^k & \text{if } I = A \\ 0 & \text{if } I \neq A. \end{cases}$$

Thus, if $I \neq A$ then $f_I dx_I(\iota) = 0$ and

$$\omega(\iota) = f_A dx_A(\iota) = r^k \int_{I^k} f_A(\iota(u))\, du.$$

Continuity of f_A implies that

$$f_A(p) = \lim_{r \to 0} \frac{1}{r^k}\, \omega(\iota), \tag{20}$$

which is how the value of ω on small k-cells at p determines the coefficient $f_A(p)$. □

38 Corollary *If $k > n$ then $\Omega^k(\mathbb{R}^n) = 0$.*

Proof There are no ascending k-tuples of integers in $\{1, , \ldots, n\}$. □

Moral A form may have many names, but it has a unique ascending name. Therefore if definitions or properties of a form are to be discussed in terms of a form's name, the use of ascending names avoids ambiguity.

Wedge Products

Let α be a k-form and β be an ℓ-form. Write them in their ascending presentations, $\alpha = \sum_I a_I dx_I$ and $\beta = \sum_J b_J dx_J$. Their **wedge product** is the $(k + \ell)$-form

$$\alpha \wedge \beta = \sum_{I,J} a_I b_J dx_{IJ}$$

where $I = (i_1, \ldots, i_k)$, $J = (j_1, \ldots, j_\ell)$, $IJ = (i_1, \ldots, i_k, j_1, \ldots, j_\ell)$, and the sum is taken over all ascending I, J. The use of ascending presentations avoids name ambiguity, although Theorem 39 makes the ambiguity moot. A particular case of the definition is

$$dx_1 \wedge dx_2 = dx_{(1,2)}.$$

39 Theorem *The wedge product $\wedge : \Omega^k \times \Omega^\ell \to \Omega^{k+\ell}$ satisfies four natural conditions:*

(a) distributivity: $(\alpha + \beta) \wedge \gamma = \alpha \wedge \gamma + \beta \wedge \gamma$ and $\gamma \wedge (\alpha + \beta) = \gamma \wedge \alpha + \gamma \wedge \beta$.

(b) insensitivity to presentations: $\alpha \wedge \beta = \sum_{I,J} a_I b_J dx_{IJ}$ for general presentations $\alpha = \sum a_I dx_I$ and $\beta = \sum b_J dx_J$.

(c) associativity: $\alpha \wedge (\beta \wedge \gamma) = (\alpha \wedge \beta) \wedge \gamma$.

(d) signed commutativity: $\beta \wedge \alpha = (-1)^{k\ell} \alpha \wedge \beta$, when α is a k-form and β is an ℓ-form. In particular, $dx \wedge dy = -dy \wedge dx$.

40 Lemma *The wedge product of basic forms satisfies*

$$dx_I \wedge dx_J = dx_{IJ}.$$

Proof #1 See Exercise 54.

Proof #2 If I and J ascend then the lemma merely repeats the definition of the wedge product. Otherwise, let π and ρ be permutations that make πI and ρJ non-descending. Call σ the permutation of IJ that is π on the first k terms and ρ on the last ℓ. The sign of σ is $\operatorname{sgn}(\pi) \operatorname{sgn}(\rho)$, and

$$dx_I \wedge dx_J = \operatorname{sgn}(\pi) \operatorname{sgn}(\rho) dx_{\pi I} \wedge dx_{\rho J} = \operatorname{sgn}(\sigma) dx_{\sigma(IJ)} = dx_{IJ}. \quad \square$$

Proof of Theorem 39 (a) To check distributivity, suppose that $\alpha = \sum a_I dx_I$ and $\beta = \sum b_I dx_I$, are k-forms, while $\gamma = \sum c_J dx_J$ is an ℓ-form, and all sums are ascending presentations. Then

$$\sum (a_I + b_I) dx_I$$

is the ascending presentation of $\alpha + \beta$ (this is the only trick in the proof), and

$$(\alpha + \beta) \wedge \gamma = \sum_{I,J} (a_I + b_I) c_J dx_{IJ} = \sum_{I,J} a_I c_J dx_{IJ} + \sum_{I,J} b_I c_J dx_{IJ},$$

which is $\alpha \wedge \gamma + \beta \wedge \gamma$, and verifies distributivity on the left. Distributivity on the right is checked in a similar way.

(b) Let $\sum a_I dx_I$ and $\sum b_J dx_J$ be general, non-ascending presentations of α and β. By distributivity and Lemma 40,

$$\Big(\sum_I a_I dx_I \Big) \wedge \Big(\sum_J b_J dx_J \Big) = \sum_{I,J} a_I b_J\, dx_I \wedge dx_J = \sum_{I,J} a_I b_J\, dx_{IJ}$$

(c) By (b), to check associativity we need not use ascending presentations. Thus, if $\alpha = \sum a_I dx_I$, $\beta = \sum b_J dx_J$, and $\gamma = \sum c_K dx_K$ then

$$\alpha \wedge (\beta \wedge \gamma) = \Big(\sum_I a_I dx_I\Big) \wedge \Big(\sum_{J,K} b_J c_K dx_{JK}\Big) = \sum_{I,J,K} a_I b_J c_K dx_{IJK},$$

which equals $(\alpha \wedge \beta) \wedge \gamma$.

(d) Associativity implies that it makes sense to write dx_I and dx_J as products $dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ and $dx_{j_1} \wedge \cdots \wedge dx_{j_\ell}$. Thus,

$$dx_I \wedge dx_J = dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_\ell}.$$

It takes $k\ell$ pair-transpositions to push each dx_i past each dx_j, which implies

$$dx_J \wedge dx_I = (-1)^{k\ell} dx_I \wedge dx_J.$$

Distributivity completes the proof of signed commutativity for the general α and β. □

The Exterior Derivative

Differentiating a form is subtle. The idea, as with all derivatives, is to imagine how the form changes under small variations of the point at which it is evaluated.

A 0-form is a smooth function $f(x)$. Its exterior derivative is by definition the functional on paths $\varphi : [0, 1] \to \mathbb{R}^n$,

$$df : \varphi \mapsto f(\varphi(1)) - f(\varphi(0)).$$

41 Proposition *df is a 1-form and when $n = 2$, it is expressed as*

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy.$$

In particular, $d(x) = dx$.

Proof When no abuse of notation occurs we use calculus shorthand and write $f_x = \partial f/\partial x$, $f_y = \partial f/\partial y$. Applied to φ, the form $\omega = f_x dx + f_y dy$ produces the number

$$\omega(\varphi) = \int_0^1 \left(f_x(\varphi(t)) \frac{dx(t)}{dt} + f_y(\varphi(t)) \frac{dy(t)}{dt} \right) dt.$$

By the Chain Rule, the integrand is the derivative of $f \circ \varphi(t)$, so the Fundamental Theorem of Calculus implies that $\omega(\varphi) = f(\varphi(1)) - f(\varphi(0))$. Therefore, $df = \omega$ as claimed. □

Definition Fix $k \geq 1$. Let $\sum f_I dx_I$ be the ascending presentation of a k-form ω. The **exterior derivative** of ω is the $(k+1)$-form

$$d\omega = \sum_I df_I \wedge dx_I.$$

The sum is taken over all ascending k-tuples I.

Use of the ascending presentation makes the definition unambiguous, although Theorem 42 makes this moot. Since df_I is a 1-form and dx_I is k-form, $d\omega$ is indeed a $(k+1)$-form.

For example we get

$$d(f dx + g dy) = (g_x - f_y) dx \wedge dy.$$

42 Theorem *Exterior differentiation $d : \Omega^k \to \Omega^{k+1}$ satisfies four natural conditions.*

(a) It is linear: $d(\alpha + c\beta) = d\alpha + c d\beta$.

(b) It is insensitive to presentation: if $\sum f_I dx_I$ is a general presentation of ω then $d\omega = \sum df_I \wedge dx_I$.

(c) It obeys a product rule: if α is a k-form and β is an ℓ-form then

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta.$$

(d) $d^2 = 0$. That is, $d(d\omega) = 0$ for all $\omega \in \Omega^k$.

Proof (a) Linearity is easy and is left for the reader as Exercise 55.

(b) Let π make πI ascending. Linearity of d and associativity of $\wedge$ give

$$d(f_I dx_I) = \operatorname{sgn}(\pi)\, d(f_I dx_{\pi I}) = \operatorname{sgn}(\pi)\, d(f_I) \wedge dx_{\pi I} = d(f_I) \wedge dx_I.$$

Linearity of d promotes the result from simple forms to general ones.

(c) The ordinary Leibniz product rule for differentiating functions of two variables gives

$$\begin{aligned} d(fg) &= \frac{\partial fg}{\partial x} dx + \frac{\partial fg}{\partial y} dy \\ &= f_x g\, dx + f_y g\, dy + f g_x\, dx + f g_y\, dy \end{aligned}$$

which is $g df + f dg$, and verifies (c) for 0-forms in $\mathbb{R}^2$. The higher-dimensional case is similar. Next we consider simple forms $\alpha = f dx_I$ and $\beta = g dx_J$. Then

$$\begin{aligned} d(\alpha \wedge \beta) &= d(fg\, dx_{IJ}) = (g df + f dg) \wedge dx_{IJ} \\ &= (df \wedge dx_I) \wedge (g dx_J) \;+\; (-1)^k (f dx_I) \wedge (dg \wedge dx_J) \\ &= d\alpha \wedge \beta + (-1)^k \alpha \wedge d\beta. \end{aligned}$$

Distributivity completes the proof for general α and β.

The proof of (d) is fun. We check it first for the special 0-form x. By Proposition 41, the exterior derivative x is dx and in turn the exterior derivative of dx is zero. For $dx = 1dx$, $d1 = 0$, and by definition, $d(1dx) = d(1) \wedge dx = 0$. For the same reason, $d(dx_I) = 0$.

Next we consider a smooth function $f : \mathbb{R}^2 \to \mathbb{R}$ and prove that $d^2 f = 0$. Since $d^2 x = d^2 y = 0$ we have

$$\begin{aligned} d^2 f &= d(f_x dx + f_y dy) = d(f_x) \wedge dx + d(f_y) \wedge dy \\ &= (f_{xx} dx + f_{xy} dy) \wedge dx + (f_{yx} dx + f_{yy} dy) \wedge dx \\ &= 0. \end{aligned}$$

The fact that $d^2 = 0$ for functions easily gives the same result for forms. The higher-dimensional case is similar. □

Pushforward and Pullback

According to Theorem 36, forms behave naturally under composition on the right. What about composition on the left? Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a smooth transformation. It induces a natural transformation $T_* : C_k(\mathbb{R}^n) \to C_k(\mathbb{R}^m)$, the **pushforward** of T, defined by

$$T_* : \varphi \mapsto T \circ \varphi.$$

Dual to the pushforward is the **pullback** $T^* : C^k(\mathbb{R}^m) \to C^k(\mathbb{R}^n)$ defined by

$$T^* : Y \mapsto Y \circ T.$$

Thus, the pullback of $Y \in C^k(\mathbb{R}^m)$ is the functional on $C_k(\mathbb{R}^n)$

$$T^* Y : \varphi \mapsto Y(\varphi \circ T).$$

The pushforward T_* goes the same direction as T, from $\mathbb{R}^n$ to $\mathbb{R}^m$, while the pullback T^* goes the opposite way. The pushforward/pullback duality is summarized by the formula

$$(T^* Y)(\varphi) = Y(T_* \varphi).$$

43 Theorem *Pullbacks obey the following four natural conditions.*

(a) The pullback transformation is linear and $(S \circ T)^* = T^* \circ S^*$.

(b) The pullback of a form is a form; in particular, $T^*(dy_I) = dT_I$ *and* $T^*(f\,dy_I) = T^* f\, dT_I$, *where* $dT_I = dT_{i_1} \wedge \cdots \wedge dT_{i_k}$.

(c) The pullback preserves wedge products, $T^*(\alpha \wedge \beta) = T^*\alpha \wedge T^*\beta$.

(d) The pullback commutes with the exterior derivative, $dT^* = T^* d$.

Proof (a) This is left as Exercise 56.

(b) We rely on a nontrivial result in linear algebra, the Cauchy-Binet Formula, which concerns the determinant of a product matrix $AB = C$, where A is $k \times n$ and B is $n \times k$. See Appendix E.

In terms of Jacobians, the Cauchy-Binet Formula asserts that if the maps $\varphi : \mathbb{R}^k \to \mathbb{R}^n$ and $\psi : \mathbb{R}^n \to \mathbb{R}^k$ are smooth, then the composite $\phi = \psi \circ \varphi : \mathbb{R}^k \to \mathbb{R}^k$ satisfies

$$\frac{\partial \phi}{\partial u} = \sum_J \frac{\partial \psi}{\partial x_J} \frac{\partial \varphi_J}{\partial u}$$

where the Jacobian $\partial\psi/\partial x_J$ is evaluated at $x = \varphi(u)$, and J ranges through all ascending k-tuples in $\{1, \ldots, n\}$.

Then the pullback of a simple k-form on $\mathbb{R}^m$ is the functional on $C_k(\mathbb{R}^n)$,

$$\begin{aligned} T^*(f\,dy_I) : \varphi &\mapsto f\,dy_I(T \circ \varphi) \\ &= \int_{I^k} f(T \circ \varphi(u)) \frac{\partial (T \circ \varphi)_I}{\partial u}\, du \\ &= \sum_J \int_{I^k} f(T \circ \varphi(u)) \Big(\frac{\partial T_I}{\partial x_J}\Big)_{x=\varphi(u)} \frac{\partial \varphi_J}{\partial u}\, du, \end{aligned}$$

which implies that

$$T^*(f\,dy_I) = \sum_J T^* f\, \frac{\partial T_I}{\partial x_J}\, dx_J$$

is a k-form. Linearity of the pullback promotes this to general forms — the pullback of a form is a form. It remains to check that $T^*(dy_I) = dT_I$. For $I = (i_1, \ldots, i_k)$, distributivity of the wedge product and the definition of the exterior derivative of a function imply that

$$\begin{aligned} dT_I = dT_{i_1} \wedge \cdots \wedge dT_{i_k} &= \Big(\sum_{s_1=1}^n \frac{\partial T_{i_1}}{\partial x_{s_1}} dx_{s_1}\Big) \wedge \cdots \wedge \Big(\sum_{s_k=1}^n \frac{\partial T_{i_k}}{\partial x_{s_k}} dx_{s_k}\Big) \\ &= \sum_{s_1, \ldots, s_k = 1}^n \frac{\partial T_{i_1}}{\partial x_{s_1}} \cdots \frac{\partial T_{i_k}}{\partial x_{s_k}}\, dx_{s_1} \wedge \cdots \wedge dx_{s_k} \end{aligned}$$

The indices $i_1, \ldots, i_k$ are fixed. All terms with repeated dummy indices $s_1, \ldots, s_k$ are zero, so the sum is really taken as $(s_1, \ldots, s_k)$ varies in the set

of k-tuples with no repeated entry, and then we know that $(s_1, \ldots, s_k)$ can be expressed uniquely as $(s_1, \ldots, s_k) = \pi J$ for an ascending $J = (j_1, \ldots, j_k)$ and a permutation π. Also, $dx_{s_1} \wedge \cdots \wedge dx_{s_k} = \operatorname{sgn}(\pi) dx_J$. This gives

$$dT_I = \sum_J \Big(\sum_\pi \operatorname{sgn}(\pi) \frac{\partial T_{i_1}}{\partial x_{\pi(j_1)}} \cdots \frac{\partial T_{i_k}}{\partial x_{\pi(j_k)}} \Big) dx_J = \sum_J \frac{\partial T_I}{\partial x_J} dx_J,$$

and hence $T^*(dy_I) = dT_I$. Here we used the description of the determinant from Appendix E.

(c) For 0-forms it is clear that the pullback of a product is the product of the pullbacks, $T^*(fg) = T^*f \, T^*g$. If the forms $\alpha = f dy_I$ and $\beta = g dy_J$ are simple then $\alpha \wedge \beta = fg dy_{IJ}$ and by (b),

$$T^*(\alpha \wedge \beta) = T^*(fg)\, dT_{IJ} = T^*f \, T^*g \, dT_I \wedge dT_J = T^*\alpha \wedge T^*\beta.$$

Wedge distributivity and pullback linearity complete the proof of (c).

(d) If ω is a form of degree 0, $\omega = f \in \Omega^0(\mathbb{R}^m)$, then

$$\begin{aligned} T^*(df)(x) &= T^*\Big(\sum_{i=1}^m \frac{\partial f}{\partial y_i} dy_i \Big) \\ &= \sum_{i=1}^m T^*\Big(\frac{\partial f}{\partial y_i} \Big) T^*(dy_i) \\ &= \sum_{i=1}^m \Big(\frac{\partial f(y)}{\partial y_i} \Big)_{y=T(x)} dT_i \\ &= \sum_{i=1}^m \sum_{j=1}^n \Big(\frac{\partial f(y)}{\partial y_i} \Big)_{y=T(x)} \Big(\frac{\partial T_i}{\partial x_j} \Big) dx_j, \end{aligned}$$

which is merely the chain rule expression for $d(f \circ T) = d(T^*f)$,

$$d(f \circ T) = \sum_{j=1}^n \Big(\frac{\partial f(T(x))}{\partial x_j} \Big) dx_j.$$

Thus, $T^*d\omega = dT^*\omega$ for 0-forms.

Next consider a simple k-form $\omega = f dy_I$ with $k \geq 1$. Using (b), the degree-zero case, and the wedge differentiation formula, we get

$$\begin{aligned} d(T^*\omega) &= d(T^*f \, dT_I) \\ &= d(T^*f) \wedge dT_I + (-1)^0 T^*f \wedge d(dT_I) \\ &= T^*(df) \wedge dT_I \\ &= T^*(df \wedge dy_I) \\ &= T^*(d\omega). \end{aligned}$$

Linearity promotes this to general k-forms and completes the proof of (d). □

10 The General Stokes' Formula

In this section we establish the general Stokes' formula as

$$\int_\varphi d\omega = \int_{\partial\varphi} \omega,$$

where $\omega \in \Omega^k(\mathbb{R}^n)$ and $\varphi \in C_{k+1}(\mathbb{R}^n)$. Then, as special cases, we reel off the standard formulas of vector calculus. Finally, we discuss antidifferentiation of forms and briefly introduce de Rham cohomology.

First we verify Stokes' formula on a cube, and then get the general case by means of the pullback.

Definition A ***k*-chain** is a formal linear combination of k-cells,

$$\Phi = \sum_{j=1}^N a_j \varphi_j,$$

where $a_1 \ldots, a_N$ are real constants. The integral of a k-form ω over Φ is

$$\int_\Phi \omega = \sum_{j=1}^N a_j \int_{\varphi_j} \omega.$$

Definition The **boundary** of a k-cell φ is the k-chain

$$\partial\varphi = \sum_{j=1}^{k+1} (-1)^{j+1} (\varphi \circ \iota^{j,1} - \varphi \circ \iota^{j,0})$$

where

$$\iota^{j,0} : (u_1, \ldots, u_k) \mapsto (u_1, \ldots, u_{j-1}, 0, u_j, \ldots, u_k)$$
$$\iota^{j,1} : (u_1, \ldots, u_k) \mapsto (u_1, \ldots, u_{j-1}, 1, u_j, \ldots, u_k).$$

are the j^{th} **rear inclusion** k-cell and j^{th} **front inclusion** k-cell of I^{k+1}. As shorthand, one can write $\partial\varphi$ as

$$\partial\varphi = \sum_{j=1}^{k+1} (-1)^{j+1} \delta^j$$

where $\delta^j = \varphi \circ \iota^{j,1} - \varphi \circ \iota^{j,0}$ is the j^{th} **dipole** of φ.

44 Stokes' Theorem for a Cube *Assume that $k+1 = n$. If $\omega \in \Omega^k(\mathbb{R}^n)$ and $\iota : I^n \to \mathbb{R}^n$ is the identity-inclusion n-cell in $\mathbb{R}^n$ then*

$$\int_\iota d\omega = \int_{\partial\iota} \omega.$$

Proof Write ω as

$$\omega = \sum_{i=1}^{n} f_i(x) dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n,$$

where the hat above the term dx_i is standard notation to indicate that it is deleted. The exterior derivative of ω is

$$\begin{aligned} d\omega &= \sum_{i=1}^{n} df_i \wedge dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n \\ &= \sum_{i=1}^{n} (-1)^{i-1} \frac{\partial f_i}{\partial x_i} dx_1 \wedge \cdots \wedge dx_n, \end{aligned}$$

which implies that

$$\int_\iota d\omega = \sum_{i=1}^{n} (-1)^{i+1} \int_{I^k} \frac{\partial f_i}{\partial x_i} du.$$

Deleting the j^{th} component of the rear j^{th} face $\iota^{j,0}(u)$ gives the k-tuple $(u_1, \ldots, u_k)$, while deleting any other component gives a k-tuple with a

component that remains constant as u varies. The same is true of the j^{th} front face. Thus the Jacobians are

$$\frac{\partial(\iota^{j,0 \text{ or } 1})_I}{\partial u} = \begin{cases} 1 & \text{if } I = (1, \ldots, \widehat{j}, \ldots, n) \\ 0 & \text{otherwise,} \end{cases}$$

and so the j^{th} dipole integral of ω is zero except when $i = j$, and in that case

$$\begin{aligned} \int_{\delta^j} \omega = \int_0^1 \cdots \int_0^1 \Big(& f_j(u_1, \ldots, u_{j-1}, 1, u_j, \ldots, u_k) \\ & - f_j(u_1, \ldots, u_{j-1}, 0, u_j, \ldots, u_k) \Big) du_1 \ldots du_k. \end{aligned}$$

By the Fundamental Theorem of Calculus we can substitute the integral of a derivative for the f_j difference; and, by Fubini's Theorem, the order of integration in ordinary multiple integration is irrelevant. This gives

$$\int_{\delta^j} \omega = \int_0^1 \cdots \int_0^1 \frac{\partial f_j(x)}{\partial x_j} dx_1 \ldots dx_n,$$

so the alternating dipole sum $\sum (-1)^{j+1} \int_{\delta^j} \omega$ equals $\int_\iota d\omega$. □

45 Corollary — Stokes' Formula for a general k-cell *If $\omega \in \Omega^k(\mathbb{R}^n)$ and if $\varphi \in C_{k+1}(\mathbb{R}^n)$ then*

$$\int_\varphi d\omega = \int_{\partial\varphi} \omega.$$

Proof Using the pullback definition and applying (d) of Theorem 43 when $T = \varphi : I^{k+1} \to \mathbb{R}^n$ and $\iota : I^{k+1} \to \mathbb{R}^{k+1}$ is the identity-inclusion gives

$$\int_\varphi d\omega = \int_{\varphi\circ\iota} d\omega = \int_\iota \varphi^* d\omega = \int_\iota d\varphi^* \omega = \int_{\partial\iota} \varphi^* \omega = \int_{\partial\varphi} \omega.$$

Stokes' Formula on manifolds

If $M \subset \mathbb{R}^n$ divides into $(k+1)$-cells and its boundary divides into k-cells, as shown in Figure 118, there is a version of Stokes' Formula for M. Namely, if ω is a k-form, then

$$\int_M d\omega = \int_{\partial M} \omega.$$

It is required that the boundary k-cells which are interior to M cancel each other out. This prohibits the Möbius band and other nonorientable sets. The $(k+1)$-cells **tile** M. Since smooth cells can be singular (not one-to-one) their images can be simplices (triangles, etc.), a helpful fact when tiling M.

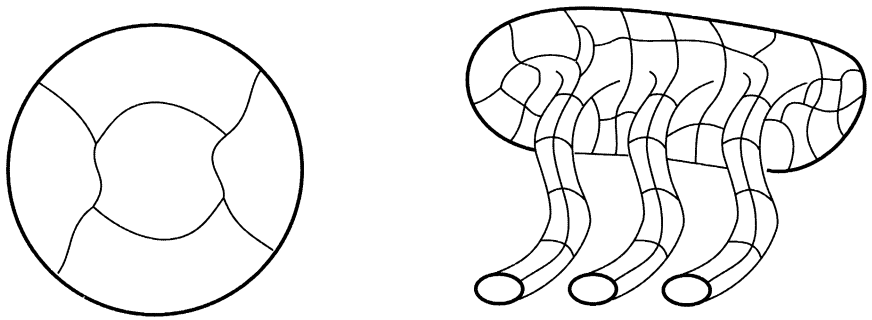


Figure 118 Manifolds of $(k+1)$-cells. The boundaries may have more than one connected component.

Vector Calculus

The Fundamental Theorem of Calculus can be viewed a special case of Stokes' Formula

$$\int_M d\omega = \int_{\partial M} \omega$$

by taking $M = [a, b] \subset \mathbb{R}^1$ and $\omega = f$. The integral of ω over the 0-chain $\partial M = b - a$ is $f(b) - f(a)$, while the integral of $d\omega$ over M is $\int_a^b f'(x)\,dx$.

Second, Green's Formula in the plane,

$$\iint_D (g_x - f_y)\,dxdy = \int_C f\,dx + g\,dy,$$

is also a special case when we take $\omega = f\,dx + g\,dy$. Here, the region D is bounded by the curve C. It is a manifold of 2-cells in the plane.

Third, the Gauss Divergence Theorem,

$$\iiint_D \operatorname{div} F = \iint_S \operatorname{flux} F,$$

is a consequence of Stokes' Formula. Here, $F = (f, g, h)$ is a smooth vector field defined on $U \subset \mathbb{R}^3$. (The notation indicates that f is the x-component of F, g is its y-component, and h is its z-component.) The **divergence** of F is the scalar function

$$\operatorname{div} F = f_x + g_y + h_z.$$

If φ is a 2-cell in U, the integral

$$\int_\varphi f\,dy \wedge dz + g\,dz \wedge dx + h\,dx \wedge dy$$

is the **flux** of F across φ. Let S be a compact manifold of 2-cells. The total flux across S is the sum of the flux across its 2-cells. If S bounds a region $D \subset U$, then the Gauss Divergence Theorem is just Stokes' Formula with

$$\omega = f\,dy \wedge dz + g\,dz \wedge dx + h\,dx \wedge dy.$$

For $d\omega = \operatorname{div} F\,dx \wedge dy \wedge dz$.

Finally, the **curl** of a vector field $F = (f, g, h)$ is the vector field

$$(h_y - g_z,\ f_z - h_x,\ g_x - f_y).$$

Applying Stokes' Formula to the form $\omega = f\,dx + g\,dy + h\,dz$ gives

$$\int_S (h_y - g_z)\,dy \wedge dz + (f_z - h_x)\,dz \wedge dx + (g_x - f_y)\,dx \wedge dy$$
$$= \int_C f\,dx + g\,dy + h\,dz$$

where S is a surface bounded by the closed curve C. The first integral is the total curl across S, while the second is the circulation of F at the boundary. Their equality is Stokes' Curl Theorem.

Closed Forms and Exact Forms

A form is **closed** if its exterior derivative is zero. It is **exact** if it is the exterior derivative of some other form. Since $d^2 = 0$, every exact form is closed:

$$\omega = d\alpha \quad \Rightarrow \quad d\omega = d(d\alpha) = 0.$$

When is the converse true? That is, when can we antidifferentiate a closed form ω? If the forms are defined on $\mathbb{R}^n$, the answer "always" is Poincaré's Lemma. See below. But if the forms are defined on some subset U of $\mathbb{R}^n$, and if they do not extend to smooth forms defined on all of $\mathbb{R}^n$, then the answer depends on the topology of U.

There is one case that should be familiar from calculus. Let U be a planar region that is bounded by a simple closed curve, and let $\omega = f\,dx + g\,dy$ be a closed 1-form on U. Then it is exact. For one can show that the integral of ω along a path $C \subset U$ depends only on the endpoints of C. The integral is "path independent." Fix a point $p \in U$ and set

$$h(q) = \int_C \omega$$

where $q \in U$ and C is any path in U from p to q. The function h is well defined because the integral depends only on the endpoints. One checks that $\partial h/\partial x = f$, $\partial h/\partial y = g$, so $dh = \omega$ and ω is exact.

An open set $U \subset \mathbb{R}^n$ is **simply connected** if every closed curve in U can be shrunk to a point in U without leaving U. A region in the plane that is bounded by a simple closed curve is simply connected. (In fact it is homeomorphic to the open disc.) Also, the n-dimensional ball is simply connected, and so is the spherical shell in $\mathbb{R}^3$ which consists of all points whose distance to the origin is between a and b with $0 < a < b$. The preceding construction of f works equally well in dimension n and implies that a 1-form defined on a simply connected region in $\mathbb{R}^n$ is closed if and only if it is exact.

If $U \subset \mathbb{R}^2$ is not simply connected, there are 1-forms on it that are closed but not exact. The standard example is

$$\omega = \frac{-y}{r^2}\,dx + \frac{x}{r^2}\,dy$$

where $r = \sqrt{x^2 + y^2}$. See Exercise 61.

In $\mathbb{R}^3$ it is instructive to consider the 2-form

$$\omega = \frac{x}{r^3}\,dy \wedge dz + \frac{y}{r^3}\,dz \wedge dx + \frac{z}{r^3}\,dx \wedge dy.$$

ω is defined on U, which is $\mathbb{R}^3$ minus the origin. U is a spherical shell with inner radius 0 and outer radius ∞. The form ω is closed but not exact despite the fact that U is simply connected.

46 Poincaré's Lemma *If ω is a closed k-form on $\mathbb{R}^n$ then it is exact.*

Proof In fact, a better result is true. There are **integration operators**

$$L_k : \Omega^k(\mathbb{R}^n) \to \Omega^{k-1}(\mathbb{R}^n)$$

with the property that $Ld + dL = \text{identity}$. That is, for all $\omega \in \Omega^k(\mathbb{R}^n)$,

$$(L_{k-1}d + dL_k)(\omega) = \omega.$$

From the existence of these integration operators, the Poincaré Lemma is immediate. For, if $d\omega = 0$ then we have

$$\omega = L(d\omega) + dL(\omega) = dL(\omega),$$

which shows that ω is exact.

The construction of L is tricky. First we consider a k-form β, not on $\mathbb{R}^n$, but on $\mathbb{R}^{n+1}$. It can be expressed uniquely as

$$\beta = \sum_I f_I\,dx_I + \sum_J g_J\,dt \wedge dx_J \tag{21}$$

where $f_I = f_I(x,t)$, $g_J = g_J(x,t)$, and $(x,t) \in \mathbb{R}^{n+1} = \mathbb{R}^n \times \mathbb{R}$. The first sum is taken over all ascending k-tuples I in $\{1,\dots,n\}$, and the second over all ascending $(k-1)$-tuples J in $\{1,\dots,n\}$. Then we define operators

$$N : \Omega^k(\mathbb{R}^{n+1}) \to \Omega^{k-1}(\mathbb{R}^n)$$

by setting

$$N(\beta) = \sum_J \Big(\int_0^1 g_J(x,t)\,dt\Big)dx_J.$$

We claim that for all $\beta \in \Omega^k(\mathbb{R}^{n+1})$,

$$(dN + Nd)(\beta) = \sum_I (f_I(x,1) - f_I(x,0))\,dx_I \tag{22}$$

where the coefficients f_I take their meaning from (21). By Theorem 14 it is legal to differentiate past the integral sign. Thus

$$d\beta = \sum_{I,\ell} \frac{\partial f_I}{\partial x_\ell} \wedge dx_I + \sum_I \frac{\partial f_I}{\partial t}\,dt \wedge dx_I + \sum_{J,\ell} \frac{\partial g_J}{\partial x_\ell}\,dx_\ell \wedge dt \wedge dx_J$$

$$N(d\beta) = \sum_I \Big(\int_0^1 \frac{\partial f_I}{\partial t}\,dt\Big)dx_I - \sum_{J,\ell} \Big(\int_0^1 \frac{\partial g_J}{\partial x_\ell}\,dt\Big)dx_\ell \wedge dx_J$$

$$dN(\beta) = \sum_{J,\ell} \Big(\int_0^1 \frac{\partial g_J}{\partial x_\ell}\,dt\Big)dx_\ell \wedge dx_J,$$

and therefore

$$(dN + Nd)(\beta) = \sum_{I,\ell} \Big(\int_0^1 \frac{\partial f_I}{\partial t}\,dt\Big)dx_I = \sum_I (f_I(x,1) - f_I(x,0))dx_I,$$

as is claimed in (22).

Then we define a "cone map" $\rho : \mathbb{R}^{n+1} \to \mathbb{R}^n$ by

$$\rho(x,t) = tx,$$

and set $L = N \circ \rho^*$. Commutativity of pullback and d gives

$$Ld + dL = N\rho^* d + dN\rho^* = (Nd + dN)\rho^*,$$

so it behooves us to work out $\rho^*(\omega)$. First suppose that ω is simple, say $\omega = h dx_I \in \Omega^k(\mathbb{R}^n)$. Since $\rho(x, t) = (tx_1, \ldots, tx_n)$, we have

$$\begin{aligned}\rho^*(h dx_I) &= (\rho^* h)(\rho^*(dx_I)) = h(tx) d\rho_I \\ &= h(tx)(d(tx_{i_1}) \wedge \cdots \wedge d(tx_{i_k})) \\ &= h(tx)((t dx_{i_1} + x_{i_1} dt) \wedge \cdots \wedge (t dx_{i_k} + x_{i_k} dt)) \\ &= h(tx)(t^k dx_I) + \text{ terms that include } dt.\end{aligned}$$

From (22) we conclude that

$$(Nd + dN) \circ \rho^*(h dx_I) = (h(1x)1^k - h(0x)0^k) dx_I = h dx_I.$$

Linearity of L and d promote this equation to general k-forms,

$$(Ld + dL)\omega = \omega,$$

and as remarked at the outset, existence of such an L implies that closed forms on $\mathbb{R}^n$ are exact. □

47 Corollary *If U is diffeomorphic to $\mathbb{R}^n$ then closed forms on U are exact.*

Proof Let $T : U \to \mathbb{R}^n$ be a diffeomorphism and assume that ω is a closed k-form on U. Set $\alpha = (T^{-1})^* \omega$. Since pullback commutes with d, α is closed on $\mathbb{R}^n$, and for some $(k-1)$-form μ on $\mathbb{R}^n$, $d\mu = \alpha$. But then

$$dT^*\mu = T^* d\mu = T^*(T^{-1})^* \omega = \omega,$$

which shows that ω is exact. □

48 Corollary *Locally, closed forms defined on an open subset of $\mathbb{R}^n$ are exact.*

Proof Locally an open subset of $\mathbb{R}^n$ is diffeomorphic to $\mathbb{R}^n$. □

49 Corollary *If $U \subset \mathbb{R}^n$ is open and starlike (in particular, if U is convex) then closed forms on U are exact.*

Proof A **starlike** set $U \subset \mathbb{R}^n$ contains a point p such that the line segment from each $q \in U$ to p lies in U. It is not hard to construct a diffeomorphism from U to $\mathbb{R}^n$. □

50 Corollary *A smooth vector field F on $\mathbb{R}^3$ (or on an open set diffeomorphic to $\mathbb{R}^3$) is the gradient of a scalar function if and only if its curl is everywhere zero.*

Proof If $F = \text{grad } f$ then

$$F = (f_x, f_y, f_z) \quad \Rightarrow \quad \text{curl } F = (f_{zy} - f_{yz},\ f_{xz} - f_{zx},\ f_{yx} - f_{xy}) = 0.$$

On the other hand, if $F = (f, g, h)$ then

$$\text{curl } F = 0 \quad \Rightarrow \quad \omega = f dx + g dy + h dz$$

is closed, and therefore exact. A function f with $df = \omega$ has gradient F. □

51 Corollary *A smooth vector field on $\mathbb{R}^3$ (or on an open set diffeomorphic to $\mathbb{R}^3$) has everywhere zero divergence if and only if it is the curl of some other vector field.*

Proof If $F = (f, g, h)$ and $G = \text{curl } F$ then

$$G = (h_y - g_z,\ f_z - h_x,\ g_x - f_y)$$

so the divergence of G is zero. On the other hand, if the divergence of $G = (A, B, C)$ is zero then the form

$$\omega = A dy \wedge dz + B dz \wedge dx + C dx \wedge dy$$

is closed, and therefore exact. If the form $\alpha = f dx + g dy + h dz$ has $d\alpha = \omega$ then $F = (f, g, h)$ has curl $F = G$. □

Cohomology

The set of exact k-forms on U is usually denoted $B^k(U)$, while the set of closed k-forms is denoted $Z^k(U)$. ("B" is for boundary and "Z" is for cycle.) Both are vector subspaces of $\Omega^k(U)$ and

$$B^k(U) \subset Z^k(U).$$

The quotient vector space

$$H^k(U) = Z^k(U)/B^k(U)$$

is the k^{th} **de Rham cohomology group** of U. Its members are the **cohomology classes** of U. As was shown above, if U is simply connected then

$H^1(U) = 0$. Also, $H^2(U) \neq 0$ when U is the three-dimensional spherical shell. If U is starlike then $H^k(U) = 0$ for all $k > 0$, and $H^0(U) = \mathbb{R}$. Cohomology necessarily reflects the *global* topology of U. For locally, closed forms are exact. The relation between the cohomology of U and its topology is the subject of algebraic topology, the basic idea being that the more complicated the set U (think of Swiss cheese), the more complicated is its cohomology and vice versa. The book *From Calculus to Cohomology* by Madsen and Tornehave provides a beautiful exposition of the subject.

11* The Brouwer Fixed Point Theorem

Let $B = B^n$ be the closed unit n-ball,

$$B = \{x \in \mathbb{R}^n : |x| \leq 1\}.$$

The following is one of the deep results in topology and analysis:

52 Brouwer's Fixed Point Theorem *If $F : B \to B$ is continuous then it has a **fixed point**, a point $p \in B$ such that $F(p) = p$.*

A relatively short proof of Brouwer's Theorem can be given using Stokes' Theorem. Note that Brouwer's Theorem is trivial when $n = 0$, for B^0 is a point and is the fixed point of F. Also, if $n = 1$, the result is a consequence of the Intermediate Value Theorem on $B^1 = [-1, 1]$. For the continuous function $F(x) - x$ is non-negative at $x = -1$ and non-positive at $x = +1$, so at some $p \in [-1, 1]$, $F(p) - p = 0$; i.e., $F(p) = p$.

The strategy of the proof in higher dimensions is to suppose that there does exist a continuous $F : B \to B$ that fails to have a fixed point, and from this supposition, to derive a contradiction, namely that the volume of B is zero. The first step in the proof is standard.

Step 1. Existence of a continuous $F : B \to B$ without a fixed point implies the existence of a smooth **retraction** T of a neighborhood U of B to ∂B. The map T sends U to ∂B and fixes every point of ∂B.

If F has no fixed point as x varies in B, then compactness of B implies that there is some $\mu > 0$ such that for all $x \in B$,

$$|F(x) - x| > \mu.$$

The Stone-Weierstrass Theorem then produces a multivariable polynomial $\widetilde{F} : \mathbb{R}^n \to \mathbb{R}^n$ that $\mu/2$-approximates F on B. The map

$$G(x) = \frac{1}{1 + \mu/2}\widetilde{F}(x)$$

is smooth and sends B into the interior of B. It μ-approximates F on B, so it too has no fixed point. The restriction of G to a small neighborhood U of B also sends U into B and has no fixed point.

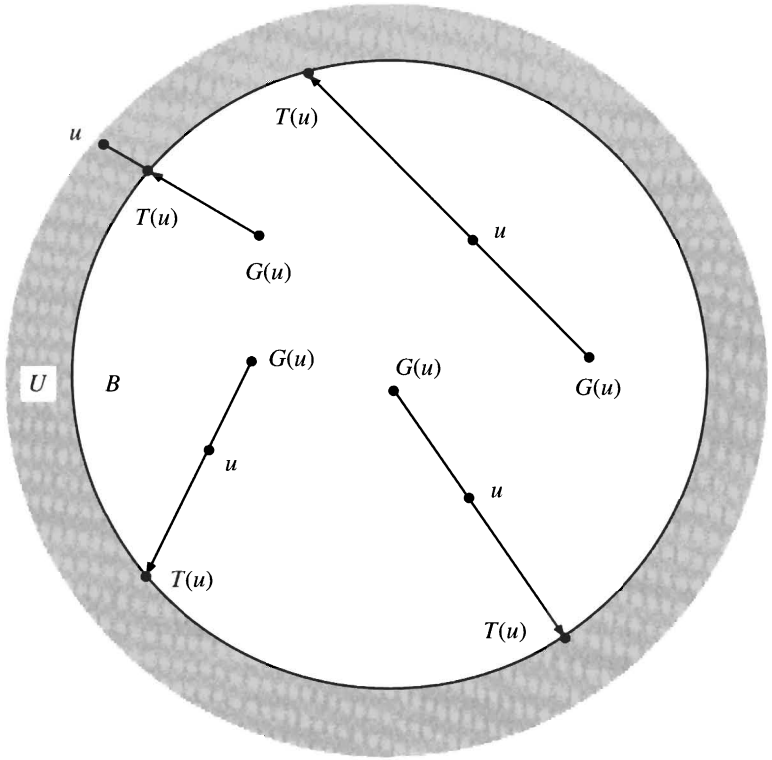


Figure 119 T retracts U onto ∂B. The point $u \in U$ is sent by T to the unique point $u' = T(u)$ at which the segment $[u, G(u)]$, extended through u, crosses the sphere ∂B.

Figure 119 shows how to construct the retraction T from the map G. Since G is smooth, so is T.

Step 2. T^* kills all n-forms. Since the range of T is ∂B, it contains no n-dimensional open set. Then the Inverse Function Theorem implies that the derivative matrix $(DT)_u$ is nowhere invertible, so its Jacobian determinant $\partial T/\partial u$ is everywhere zero, and $T^* : \Omega^n(\mathbb{R}^n) \to \Omega^n(U)$ is the zero map.

Step 3. There is a map $\varphi : I^n \to B$ that exhibits B as an n-cell such that

(a) φ is smooth.

(b) $\varphi(I^n) = B$ and $\varphi(\partial I^n) = \partial B$.

(c) $\displaystyle\int_{I^n} \frac{\partial \varphi}{\partial u}\, du > 0.$

To construct φ, start with a smooth function $\sigma : \mathbb{R} \to \mathbb{R}$ such that $\sigma(r) = 0$ for $r \leq 1/2$, $\sigma'(r) > 0$ for $1/2 < r < 1$, and $\sigma(r) = 1$ for $r \geq 1$. Then define $\psi : [-1, 1]^n \to \mathbb{R}^n$ by

$$\psi(v) = \begin{cases} v + \sigma(|v|)\left(\dfrac{v}{|v|} - v\right) & \text{if } v \neq 0 \\ 0 & \text{if } v = 0. \end{cases}$$

In fact, this formula defines ψ on all of $\mathbb{R}^n$ and ψ carries the sphere S_r of radius r to the sphere of radius

$$\rho(r) = r + \sigma(r)(1 - r),$$

sending each radial line to itself. Set $\varphi = \psi \circ \kappa$ where κ scales I^n to $[-1, 1]^n$ by the affine map $u \mapsto v = (2u_1 - 1, \ldots, 2u_n - 1)$. Then

(a) φ is smooth because, away from $u_0 = (1/2, \ldots, 1/2)$, it is the composite of smooth functions, and in a neighborhood of u_0 it is identically equal to $\kappa(u)$.

(b) Since $0 \leq \rho(r) \leq 1$ for all r, ψ sends $\mathbb{R}^n$ to B. Since $\rho(r) = 1$ for all $r \geq 1$, φ sends ∂I^n to ∂B.

(c) It is left as Exercise 65 to show that the Jacobian of ψ for $r = |v|$ is $\rho'(r)\rho(r)^{n-1}/r^{n-1}$. Thus, the Jacobian $\partial\varphi/\partial u$ is always nonnegative, and is identically equal to 2^n on the ball of radius $1/4$ at u_0, so its integral on I^n is positive.

Step 4. Consider an $(n-1)$-form α. If $\beta : I^{n-1} \to \mathbb{R}^n$ is an $(n-1)$-cell whose image lies in ∂B then

$$\int_\beta \alpha = \int_{T\circ\beta} \alpha = \int_\beta T^*\alpha.$$

The $(n-1)$-dimensional faces of $\varphi : I^n \to B$ lie in ∂B. Thus

$$\int_{\partial\varphi} \alpha = \int_{\partial\varphi} T^*\alpha. \tag{23}$$

Step 5. Now we get the contradiction. Consider the specific $(n-1)$-form

$$\alpha = x_1\, dx_2 \wedge \cdots \wedge dx_n.$$

Note that $d\alpha = dx_1 \wedge \cdots \wedge dx_n$ is n-dimensional volume and

$$\int_\varphi d\alpha = \int_{I^n} \frac{\partial\varphi}{\partial u}\, du > 0.$$

In fact the integral is the volume of B. However, we also have

$$\begin{aligned} \int_\varphi d\alpha &= \int_{\partial\varphi} \alpha && \text{by Stokes' Theorem on a cell} \\ &= \int_{\partial\varphi} T^*\alpha && \text{by Equation (23)} \\ &= \int_\varphi dT^*\alpha && \text{by Stokes' Theorem on a cell} \\ &= \int_\varphi T^* d\alpha && \text{by (d) in Theorem 43} \\ &= 0 && \text{by Step 2.} \end{aligned}$$

This is a contradiction: an integral can not simultaneously be zero and positive. The assumption that there exists a continuous $F : B \to B$ with no fixed point has led to a contradiction. Therefore it is untenable, and every F does have a fixed point.

Appendix A: Perorations of Dieudonné

In his classic book, *Foundations of Analysis*, Jean Dieudonné of the French Bourbaki school writes

> "The subject matter of this Chapter [Chapter VIII on differential calculus] is nothing else but the elementary theorems of Calculus, which however are presented in a way which will probably be new to most students. That presentation which throughout adheres strictly to our general 'geometric' outlook on Analysis, aims at keeping as close as possible to the fundamental idea of Calculus, namely the local approximation of functions by linear functions. In the classical teaching of Calculus, this idea is immediately obscured by the accidental fact that, on a one-dimensional vector space, there is a one-to-one correspondence between linear forms and numbers, and therefore the derivative at a point is defined as a number instead of a linear form. This slavish subservience to the shibboleth of numerical interpretation at any cost becomes much worse when dealing with functions of several variables: one thus arrives, for instance, at the classical formula" ..."giving the partial derivatives of a composite function, which has lost any trace of intuitive meaning, whereas the natural statement of the theorem is of course that the (total) derivative of a composite function is the composite of their derivatives" ..., "a

very sensible formulation when one thinks in terms of linear approximation."

"This 'intrinsic' formulation of Calculus, due to its greater 'abstraction', and in particular to the fact that again and again, one has to leave the initial spaces and climb higher and higher to new 'function spaces' (especially when dealing with the theory of higher derivatives), certainly requires some mental effort, contrasting with the comfortable routine of the classical formulas. But we believe the result is well worth the labor, as it will prepare the student to the still more general idea of Calculus on a differentiable manifold; the reader who wants to have a glimpse of that theory and of the questions to which it leads can look into the books of Chevalley and de Rham. Of course, he will observe in these applications, all the vector spaces which intervene have finite dimension; if that gives him an additional feeling of security, he may of course add that assumption to all the theorems of this chapter. But he will inevitably realize that this does not make the proofs shorter or simpler by a single line; in other words the hypothesis of finite dimension is entirely irrelevant to the material developed below; we have therefore thought it best to dispense with it altogether, although the applications of Calculus which deal with the finite dimensional case still by far exceed the others in number and importance."

I share most of Dieudonné's opinions expressed here. And where else will you read the phrase "slavish subservience to the shibboleth of numerical interpretation at any cost"?

Appendix B: The History of Cavalieri's Principle

The following is from Marsden and Weinstein's *Calculus*.

The idea behind the slice method goes back, beyond the invention of calculus, to Francesco Bonaventura Cavalieri (1598 - 1647), a student of Galileo and then professor at the University of Bologna. An accurate report of the events leading to Cavalieri's discovery is not available, so we have taken the liberty of inventing one.

Cavalieri's delicatessen usually produced bologna in cylindrical form, so that the volume would be computed as $\pi \cdot \text{radius}^2 \cdot \text{length}$. One day the casings were a bit weak, and

the bologna came out with odd bulges. The scale was not working that day, either, so the only way to compute the price of the bologna was in terms of its volume.

Cavalieri took his best knife and sliced the bologna into n very thin slices, each of thickness x, and measured the radii, $r_1, r_2, \ldots, r_n$ of the slices (fortunately they were all round). He then estimated the volume to be $\sum_{i=1}^{n} \pi r_i^2 x$, the sum of the volumes of the slices.

Cavalieri was moonlighting from his regular job as a professor at the University of Bologna. That afternoon he went back to his desk and began the book *Geometria indivisibilium continuorum nova quandum ratione promota* (Geometry shows the continuous indivisibility between new rations and getting promoted), in which he stated what is now known as Cavalieri's principle: If two solids are sliced by a family of parallel planes in such a way that corresponding sections have equal areas, then the two solids have the same volume.

The book was such a success that Cavalieri sold his delicatessen and retired to a life of occasional teaching and eternal glory.

Appendix C: A Short Excursion into the Complex Field

The field $\mathbb{C}$ of complex numbers corresponds bijectively with $\mathbb{R}^2$. The complex number $z = x + iy \in \mathbb{C}$ corresponds to $(x, y) \in \mathbb{R}^2$. A function $T : \mathbb{C} \to \mathbb{C}$ is **complex linear** if for all $\lambda, z, w \in \mathbb{C}$,

$$T(z + w) = T(z) + T(w) \quad \text{and} \quad T(\lambda z) = \lambda T(z).$$

Since $\mathbb{C}$ is a one-dimensional complex vector space the value $\mu = T(1)$ determines T; namely, $T(z) \equiv \mu z$. If $z = x + iy$ and $\mu = \alpha + i\beta$ then $\mu z = (\alpha x - \beta y) + i(\beta x + \alpha y)$. In $\mathbb{R}^2$ terms, $T : (x, y) \mapsto ((\alpha x - \beta y), (\beta x + \alpha y))$ which shows that T is a linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$, whose matrix is

$$\begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}.$$

The form of this matrix is special. For instance it could never be $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$.

A complex function of a complex variable $f(z)$ has a **complex derivative** $f'(z)$ if the complex ratio $(f(z+h) - f(z))/h$ tends to $f'(z)$ as the complex

number h tends to zero. Equivalently,

$$\frac{f(z+h) - f(z) - f'(z)h}{h} \to 0$$

as $h \to 0$. Write $f(z) = u(x, y) + iv(x, y)$ where $z = x + iy$, and u, v are real valued functions. Define $F : \mathbb{R}^2 \to \mathbb{R}^2$ by $F(x, y) = (u(x, y), v(x, y))$. Then F is $\mathbb{R}$-differentiable with derivative matrix

$$DF = \begin{bmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{bmatrix}.$$

Since this derivative matrix is the $\mathbb{R}^2$ expression for multiplication by the complex number $f'(z)$, it must have the $\begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}$ form. This demonstrates a basic fact about complex differentiable functions — their real and imaginary parts, u and v, satisfy the

53 Cauchy-Riemann Equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \textit{and} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Appendix D: Polar Form

The shape of the image of a unit ball under a linear transformation T is not an issue that is used directly in anything we do in Chapter 5, but it certainly underlies the geometric outlook on linear algebra.
Question. What shape is the $(n-1)$-sphere S^{n-1}?
Answer. Round.
Question. What shape is $T(S^{n-1})$?
Answer. Ellipsoidal.

Let $z = x + iy$ be a nonzero complex number. Its polar form is $z = re^{i\theta}$ where $r > 0$ and $0 \le \theta < 2\pi$, and $x = r\cos\theta$, $y = r\sin\theta$. Multiplication by z breaks up into multiplication by r, which is just dilation, and multiplication by $e^{i\theta}$, which is rotation of the plane by angle θ. As a matrix the rotation is

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

The polar coordinates of (x, y) are (r, θ).

Analogously, consider an isomorphism $T : \mathbb{R}^n \to \mathbb{R}^n$. Its **polar form** is

$$T = OP$$

where O and P are isomorphisms $\mathbb{R}^n \to \mathbb{R}^n$ such that
(a) O is like $e^{i\theta}$; it is an orthogonal isomorphism.
(b) P is like r; it is positive definite symmetric (PDS) isomorphism.

Orthogonality of O means that for all $v, w \in \mathbb{R}^n$,

$$\langle Ov, Ow \rangle = \langle v, w \rangle,$$

while P being **positive definite symmetric** means that for all nonzero vectors $v, w \in \mathbb{R}^n$,

$$\langle Pv, v \rangle > 0 \quad \text{and} \quad \langle Pv, w \rangle = \langle v, Pw \rangle.$$

The notation $\langle v, w \rangle$ indicates the usual dot product.

The polar form $T = OP$ reveals everything geometric about T. The geometric effect of O is nothing. It is an isometry and changes no distances or shapes. It is rigid. The effect of a PDS operator P is easy to describe. In linear algebra it is shown that there exists a basis $\mathcal{B} = \{u_1, \ldots, u_n\}$ of orthonormal vectors (the vectors are of unit length and are mutually perpendicular) and with respect to this basis

$$P = \begin{bmatrix} \lambda_1 & 0 & \ldots & & \\ 0 & \lambda_2 & 0 & \ldots & \\ & & \ldots & & \\ & \ldots & 0 & \lambda_{n-1} & 0 \\ & & \ldots & 0 & \lambda_n \end{bmatrix}$$

The diagonal entries λ_i are positive. P stretches each u_i by the factor λ_i. Thus P stretches the unit sphere to an n-dimensional ellipsoid. The u_i are its axes. The norm of P and hence of T is the largest λ_i, while the conorm is the smallest λ_i. The ratio of the largest to the smallest, the **condition number**, is the eccentricity of the ellipsoid.

Upshot Except for the harmless orthogonal factor O, an isomorphism is no more geometrically complicated than a diagonal matrix with positive entries.

54 Polar Form Theorem *Any isomorphism $T : \mathbb{R}^n \to \mathbb{R}^n$ factors as $T = OP$ where O is orthogonal and P is PDS.*

Proof Recall that the transpose of $T : \mathbb{R}^n \to \mathbb{R}^n$ is the unique isomorphism T^t satisfying the equation

$$\langle Tv, w\rangle = \langle v, T^t w\rangle$$

for all $v, w \in \mathbb{R}^n$. Thus, the condition $\langle Pv, w\rangle = \langle v, Pw\rangle$ in the definition of PDS means exactly that $P^t = P$.

Let T be a given isomorphism $T : \mathbb{R}^n \to \mathbb{R}^n$. We must find its factors O and P. We just write them down as follows. Consider the composite $T^t \circ T$. It is PDS because

$$(T^tT)^t = (T^t)(T^t)^t = T^tT \quad \text{and} \quad \langle T^tTv, v\rangle = \langle Tv, Tv\rangle > 0.$$

Every PDS transformation has a unique PDS square root, just as does every positive real number r. (To see this, take the diagonal matrix with entries $\sqrt{\lambda_i}$ in place of λ_i.) Thus, T^tT has a PDS square root and this is the factor P that we seek,

$$P^2 = T^tT.$$

By P^2 we mean the composite $P \circ P$. In order for the formula $T = OP$ to hold with this choice of P, we must have $O = TP^{-1}$. To finish the proof we merely must check that TP^{-1} actually is orthogonal. Magically,

$$\begin{aligned}\langle Ov, Ow\rangle &= \langle TP^{-1}v, TP^{-1}w\rangle = \langle P^{-1}v, T^tTP^{-1}w\rangle \\ &= \langle P^{-1}v, Pw\rangle = \langle P^tP^{-1}v, w\rangle = \langle PP^{-1}v, w\rangle \\ &= \langle v, w\rangle.\end{aligned}$$ □

55 Corollary *Under any invertible $T : \mathbb{R}^n \to \mathbb{R}^n$, the unit ball is sent to an ellipsoid.*

Proof Write T in polar form $T = OP$. The image of the unit ball under P is an ellipsoid. The orthogonal factor O merely rotates the ellipsoid. □

Appendix E: Determinants

A **permutation** of a set S is a bijection $\pi : S \to S$. That is, π is one-to-one and onto. We assume the set S is finite, $S = \{1, , \ldots k\}$. The **sign** of π is

$$\operatorname{sgn}(\pi) = (-1)^r$$

where r is the number of reversals — i.e., the number of pairs ij such that

$$i < j \text{ and } \pi(i) > \pi(j).$$

Proposition *Any permutation is the composite of pair transpositions; the sign of a composite permutation is the product of the signs of its factors; and the sign of a pair transposition is -1.*

The proof of this combinatorial proposition is left to the reader. Although the factorization of a permutation π into pair transpositions is not unique, the number of factors, say t, satisfies $(-1)^t = \operatorname{sgn}(\pi)$.

Definition The **determinant** of a $k \times k$ matrix A is the sum

$$\det A = \sum_{\pi} \operatorname{sgn}(\pi) a_{1\pi(1)} a_{2\pi(2)} \cdots a_{k\pi(k)}$$

where π ranges through all permutations of $\{1, \ldots, k\}$.

Equivalent definitions appear in standard linear algebra courses. One of the key facts about determinants is the product rule: for $k \times k$ matrices,

$$\det AB = \det A \det B.$$

It extends to non-square matrices as follows.

56 Cauchy-Binet Formula *Assume that $k \le n$. If A is a $k \times n$ matrix and B is an $n \times k$ matrix, then the determinant of the product $k \times k$ matrix $AB = C$ is given by the formula*

$$\det C = \sum_{J} \det A^J \det B_J$$

where J ranges through the set of ascending k-tuples in $\{1, \ldots, n\}$, A^J is the $k \times k$ minor of A whose column indices j belong to J, while B_J is the $k \times k$ minor of B whose row indices i belong to J. See Figure 120.

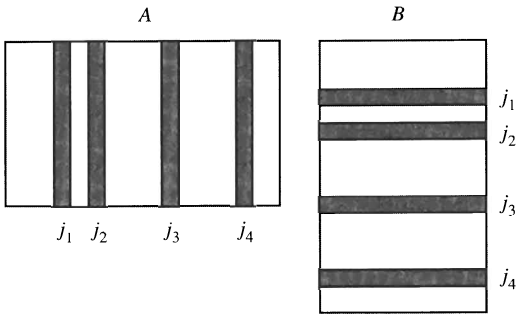


Figure 120 The paired 4×4 minors of A and B are determined by the 4-tuple $J = (j_1, j_2, j_3, j_4)$.

Proof Note that special cases of the Cauchy-Binet formula occur when $k = 1$ or $k = n$. When $k = 1$, C is the 1×1 matrix that is the dot product of an A row vector of length n times a B column vector of height n. The 1-tuples J in $\{1, \ldots, n\}$ are just single integers, $J = (1), \ldots, J = (n)$, and the product formula is immediate. In the second case, $k = n$, we have the usual product determinant formula because there is only one ascending k-tuple in $\{1, \ldots k\}$, namely $J = (1, \ldots, k)$.

To handle the general case, define the sum

$$S(A, B) = \sum_J \det A^J \det B_J$$

as above. Consider an elementary $n \times n$ matrix E. We claim that

$$S(A, B) = S(AE, E^{-1}B).$$

Since there are only two types of elementary matrices, this is not too hard a calculation, and is left to the reader. Then we perform a sequence of elementary column operations on A to put it in lower triangular form

$$A' = AE_1 \ldots E_r = \begin{bmatrix} \alpha_{11} & 0 & \ldots & \ldots & 0 & \ldots & 0 \\ \alpha_{21} & \alpha_{22} & \ldots & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & & \vdots & & \vdots \\ \alpha_{k1} & \alpha_{k2} & \ldots & \alpha_{kk} & 0 & \ldots & 0 \end{bmatrix}.$$

About $B' = E_r^{-1} \ldots E_1^{-1} B$ we observe only that

$$AB = A'B' = A'^{J_0} B'_{J_0}$$

where $J_0 = (1, \ldots, k)$. Since elementary column operations do not affect S,

$$S(A, B) = S(AE_1, E_1^{-1}B) = S(AE_1E_2, E_2^{-1}E_1^{-1}B) = \cdots = S(A', B').$$

All terms in the sum that defines $S(A', B')$ are zero except the J_0^{th}, and thus

$$\det(AB) = \det A'^{J_0} \det B'_{J_0} = S(A', B') = S(A, B)$$

as claimed. □

Exercises

1. Let $T : V \to W$ be a linear transformation, and let $p \in V$ be given. Prove that the following are equivalent.
   (a) T is continuous at the origin.
   (b) T is continuous at p.
   (c) T is continuous at at least one point of V.
2. Let $\mathcal{L}$ be the vector space of continuous linear transformations from a normed space V to a normed space W. Show that the operator norm makes $\mathcal{L}$ a normed space.
3. Let $T : V \to W$ be a linear transformation between normed spaces. Show that

$$\begin{aligned} \|T\| &= \sup\{|Tv| : |v| < 1\} \\ &= \sup\{|Tv| : |v| \leq 1\} \\ &= \sup\{|Tv| : |v| = 1\} \\ &= \inf\{M : v \in V \Rightarrow |Tv| \leq M\,|v|\}. \end{aligned}$$

4. The **conorm** of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is

$$\text{conorm}(T) = \inf\{\frac{|Tv|}{|v|} : v \neq 0\}.$$

   It is the **minimum stretch** that T imparts to vectors in $\mathbb{R}^n$. Let U be the unit ball in $\mathbb{R}^n$.
   (a) Show that the norm and conorm of T are the radii of the smallest ball that contains TU and the largest ball contained in TU.
   (b) Is the same true in normed spaces?
   (c) If T is an isomorphism, prove that its conorm is positive.
   (d) Is the converse to (c) true?
   (e) If $T : \mathbb{R}^n \to \mathbb{R}^n$ has positive conorm, why is T is an isomorphism?
   (f) If the norm and conorm of T are equal, what can you say about T?
5. Formulate and prove the fact that function composition is associative. Why can you infer that matrix multiplication is associative?
6. Let $\mathcal{M}_n$ and $\mathcal{L}_n$ be the vector spaces of $n \times n$ matrices and linear transformations $\mathbb{R}^n \to \mathbb{R}^n$.
   (a) Look up the definition of "ring" in your algebra book.
   (b) Show that $\mathcal{M}_n$ and $\mathcal{L}_n$ are rings with respect to matrix multiplication and composition.

(c) Show that $\mathcal{T} : \mathcal{M}_n \to \mathcal{L}_n$ is a ring isomorphism.

7. Two norms, $|\ \ |_1$ and $|\ \ |_2$, on a vector space are **comparable**[†] if there are positive constants c, C such that for all nonzero vectors in V,

$$c \leq \frac{|v|_1}{|v|_2} \leq C.$$

(a) Prove that comparability is an equivalence relation on norms.
(b) Prove that any two norms on a finite-dimensional vector space are comparable. [Hint: Use Theorem 2.]
(c) Consider the norms

$$|f|_{L^1} = \int_0^1 |f(t)|\, dt \quad \text{and} \quad |f|_{C^0} = \max\{|f(t)| : t \in [0, 1]\},$$

defined on the infinite-dimensional vector space C^0 of continuous functions $f : [0, 1] \to \mathbb{R}$. Show that the norms are not comparable by finding functions $f \in C^0$ whose integral norm is small but whose C^0 norm is 1.

*8. Let $|\ \ | = |\ \ |_{C^0}$ be the supremum norm on C^0 as in the previous exercise. Define an integral transformation $T : C^0 \to C^0$ by

$$T(f)(x) = \int_0^x f(t)\, dt.$$

(a) Show that T is continuous and find its norm.
(b) Let $f_n(t) = \cos(nt)$, $n = 1, 2, \ldots$. What is $T(f_n)$?
(c) Is the set of functions $K = \{T(f_n) : n \in \mathbb{N}\}$ closed? bounded? compact?
(d) Is $T(K)$ compact? How about its closure?

9. Give an example of two 2×2 matrices such that the norm of the product is less than the product of the norms.

10. In the proof of Theorem 2 we used the fact that with respect to the Euclidean norm, the length of a vector is at least as large as the length of any of its components. Show by example that this false for some norms in $\mathbb{R}^2$. [Hint: Consider the matrix

$$A = \begin{bmatrix} 3 & -2 \\ -2 & 2 \end{bmatrix}.$$

[†] From an analyst's point of view, the choice between comparable norms has little importance. At worst it affects a few constants that turn up in estimates.

Use A to define an inner product $\langle v, w\rangle_A = \sum v_i a_{ij} w_j$ on $\mathbb{R}^2$, and use the inner product to define a norm

$$|v|_A = \sqrt{\langle v, v\rangle_A}.$$

(What properties must A have for the sum to define an inner product? Does A have these properties?) With respect to this norm, what are the lengths of e_1, e_2, and $v = e_1 + e_2$?]

11. Consider the shear matrix

$$S = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$$

and the linear transformation $S : \mathbb{R}^2 \to \mathbb{R}^2$ it represents. Calculate the norm and conorm of S. [Hint: Using polar form, it suffices to calculate the norm and conorm of the positive definite symmetric part of S. Recall from linear algebra that the eigenvalues of the square of a matrix A are the squares of the eigenvalues of A.]

12. What is the one line proof that if V is a finite-dimensional normed space then its unit sphere $\{v : |v| = 1\}$ is compact?

13. The set of invertible $n \times n$ matrices is open in $\mathcal{M}$. Is it dense?

14. An $n \times n$ matrix is diagonalizable if there is a change of basis in which it becomes diagonal.
(a) Is the set of diagonalizable matrices open in $\mathcal{M}(n \times n)$?
(b) closed?
(c) dense?

15. Show that both partial derivatives of the function

$$f(x, y) = \begin{cases} \dfrac{xy}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$$

exist at the origin, but the function is not differentiable there.

16. Let $f : \mathbb{R}^2 \to \mathbb{R}^3$ and $g : \mathbb{R}^3 \to \mathbb{R}$ be defined by $f = (x, y, z)$ and $g = w$ where

$$w = w(x, y, z) = xy + yz + zx$$
$$x = x(s, t) = st \quad y = y(s, t) = s\cos t \quad z = z(s, t) = s\sin t.$$

(a) Find the matrices that represent the linear transformations $(Df)_p$ and $(Dg)_q$ where $p = (s_0, t_0) = (0, 1)$ and $q = f(p)$.
(b) Use the Chain Rule to calculate the 1×2 matrix $[\partial w/\partial s, \partial w/\partial t]$ that represents $(D(g \circ f))_p$.

(c) Plug the functions $x = x(s,t)$, $y = y(s,t)$, and $z = z(s,t)$ directly into $w = w(x,y,z)$, and recalculate $[\partial w/\partial s, \partial w/\partial t]$, verifying the answer given in (b).

(d) Examine the statements of the multivariable Chain Rules that appear in your old calculus book and observe that they are nothing more than the components of various product matrices.

17. Let $f : U \to \mathbb{R}^m$ be differentiable, $[p,q] \subset U \subset \mathbb{R}^n$, and ask whether the direct generalization of the one-dimensional Mean Value Theorem is true: does there exist a point $\theta \in [p,q]$ such that

$$f(q) - f(p) = (Df)_\theta(q-p)? \tag{24}$$

(a) Take $n = 1$, $m = 2$, and examine the function

$$f(t) = (\cos t, \sin t)$$

for $\pi \le t \le 2\pi$. Take $p = \pi$ and $q = 2\pi$. Show that there is no $\theta \in [p,q]$ which satisfies (24).

(b) Assume that the set of derivatives

$$\{(Df)_x \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) : x \in [p,q]\}$$

is convex. Prove that there exists $\theta \in [p,q]$ which satisfies (24).

(c) How does (b) imply the one dimensional Mean Value Theorem?

18. The **directional derivative** of $f : U \to \mathbb{R}^m$ at $p \in U$ in the direction u is the limit, if it exists,

$$\nabla_p f(u) = \lim_{t\to 0} \frac{f(p+tu) - f(p)}{t}.$$

(Usually, one requires that $|u| = 1$.)

(a) If f is differentiable at p, why is it obvious that the directional derivative exists in each direction u?

(b) Show that the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) = \begin{cases} \dfrac{x^3 y}{x^4 + y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

has $\nabla_{(0,0)} f(u) = 0$ for all u but is not differentiable at $(0,0)$.

*19. Using the functions in Exercises 15 and 18, show that the composite of functions whose partial derivatives exist may fail to have partial derivatives, and the composite of functions whose directional derivatives exist may fail to have directional derivatives. (That is, the classes of these functions are not closed under composition, which is further reason to define multidimensional differentiability in terms of Taylor approximation, and not in terms of partial or directional derivatives.)

20. Assume that U is a connected open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}^m$ is differentiable everywhere on U. If $(Df)_p = 0$ for all $p \in U$, show that f is constant.

21. For U as above, assume that f is second-differentiable everywhere and $(D^2 f)_p = 0$ for all p. What can you say about f? Generalize to higher-order differentiability.

22. If Y is a metric space and $f : [a,b] \times Y \to \mathbb{R}$ is continuous, show that

$$F(y) = \int_a^b f(x,y)\,dx$$

is continuous.

23. Assume that $f : [a,b] \times Y \to \mathbb{R}^m$ is continuous, Y is an open subset of $\mathbb{R}^n$, the the partial derivatives $\partial f_i(x,y)/\partial y_j$ exist, and they are continuous. Let $D_y f$ be the linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ which is represented by the matrix of partials.

(a) Show that

$$F(y) = \int_a^b f(x,y)\,dx$$

is of class C^1 and

$$(DF)_y = \int_a^b (D_y f)\,dx.$$

This generalizes Theorem 14 to higher dimensions.

(b) Generalize (a) to higher order differentiability.

24. Show that all second partial derivatives of the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) = \begin{cases} \dfrac{xy(x^2 - y^2)}{x^2 + y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

exist everywhere, but the mixed second partials are unequal at the origin, $\partial^2 f(0,0)/\partial x \partial y \neq \partial^2 f(0,0)/\partial y \partial x$.

*25. Construct an example of a C^1 function $f : \mathbb{R} \to \mathbb{R}$ that is second-differentiable only at the origin. (Infer that this phenomenon occurs also in higher dimensions.)

26. Suppose that $u \mapsto \beta_u$ is a continuous function from $U \subset \mathbb{R}^n$ into $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.
   (a) If for some $p \in U$, β_p is not symmetric, prove that its average over some small 2-dimensional parallelogram at p is also not symmetric.
   (b) Generalize (a) by replacing $\mathcal{L}$ with a finite dimensional space E, and the subset of symmetric bilinear maps with a linear subspace of E. If the average values of a continuous function always lie in the subspace then the values do too.

*27. Assume that $f : U \to \mathbb{R}^m$ is of class C^2 and show that $D^2 f$ is symmetric by the following integral method. With reference to the signed sum Δ of f at the vertices of the parallelogram P in Figure 105, use the C^1 Mean Value Theorem to show that

$$\Delta = \int_0^1 \int_0^1 (D^2 f)_{p+sv+tw}\, ds dt\, (v, w).$$

Infer symmetry of $(D^2 f)_p$ from symmetry of Δ and Exercise 26.

28. Let $\beta : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}^m$ be r-linear. Define its **symmetrization** as

$$\operatorname{symm}(\beta)(v_1, \ldots, v_r) = \frac{1}{r!} \sum_{\sigma \in P(r)} \beta(v_{\sigma(1)}, \ldots, v_{\sigma(r)}),$$

where $P(r)$ is the set of permutations of $\{1, \ldots, r\}$.
   (a) Prove that $\operatorname{symm}(\beta)$ is symmetric.
   (b) If β is symmetric prove that $\operatorname{symm}(\beta) = \beta$.
   (c) Is the converse to (b) true?
   (d) Prove that $\alpha = \beta - \operatorname{symm}(\beta)$ is antisymmetric in the sense that if σ is any permutation of $\{1, \ldots, r\}$ then

$$\alpha(v_{\sigma(1)}, \ldots, v_{\sigma(r)}) = \operatorname{sign}(\sigma)\alpha(v_1, \ldots, v_r).$$

   Infer that $\mathcal{L}^r = \mathcal{L}^r_s \oplus \mathcal{L}^r_a$ where $\mathcal{L}^r_s$ and $\mathcal{L}^r_a$ are the subspaces of symmetric and antisymmetric r-linear transformations.
   (e) Let $\beta \in \mathcal{L}^2(\mathbb{R}^2, \mathbb{R})$ be defined by

$$\beta((x, y), (x', y')) = xy'.$$

   Express β as the sum of a symmetric and an antisymmetric bilinear transformation.

*29. Prove Corollary 18 that r^{th} order differentiability implies symmetry of $D^r f$, $r \geq 3$, in one of two ways.
   (a) Use induction to show that $(D^r f)_p(v_1 \ldots, v_r)$ is symmetric with respect to permutations of $v_1, \ldots, v_{r-1}$ and of $v_2, \ldots, v_r$. Then take advantage of the fact that r is strictly greater than 2.
   (b) Define the signed sum Δ of f at the vertices of the parallelotope P spanned by $v_1, \ldots, v_r$, and show that it is the average of $D^r f$. Then proceed as in Exercise 27.

30. Consider the equation

$$xe^y + ye^x = 0. \tag{25}$$

   (a) Observe that there is no way to write down an explicit solution $y = y(x)$ of (25) in a neighborhood of the point $(x_0, y_0) = (0, 0)$.
   (b) Why, nevertheless, does there exist a C^∞ solution $y = y(x)$ of (25) near $(0, 0)$?
   (c) What is its derivative at $x = 0$?
   (d) What is its second derivative at $x = 0$?
   (e) What does this tell you about the graph of the solution?
   (f) Do you see the point of the Implicit Function Theorem better?

**31. Consider a function $f : U \to \mathbb{R}$ such that
   (i) U is a connected open subset of $\mathbb{R}^2$.
   (ii) f is C^1.
   (iii) For all $(x, y) \in U$,

$$\frac{\partial f(x, y)}{\partial y} = 0.$$

   (a) If U is a disc, show that f is independent from y.
   (b) Construct such an f of class C^∞ which does depend on y.
   (c) Show that the f in (b) can not be analytic.
   (d) Why does your example in (b) not invalidate the proof of the Rank Theorem on page 294?

32. Let G denote the set of invertible $n \times n$ matrices.
   (a) Prove that G is an open subset of $\mathcal{M}(n \times n)$.
   (b) Prove that G is a group. (It is called the **general linear group**.)
   (c) Prove that the inversion operator $\operatorname{Inv} : A \mapsto A^{-1}$ is a homeomorphism of G onto G.
   (d) Prove that Inv is a diffeomorphism and show that its derivative at A is the linear transformation $\mathcal{M} \to \mathcal{M}$,

$$X \mapsto -A^{-1} \circ X \circ A^{-1}.$$

Relate this formula to the ordinary derivative of $1/x$ at $x = a$.

33. Observe that $Y = \text{Inv}(X)$ is solves the implicit function problem

$$F(X, Y) - I = 0,$$

where $F(X, Y) = X \circ Y$. Assume it is known that Inv is smooth and use the chain rule to derive from this equation the formula for the derivative of Inv.

34. Use Gaussian elimination to prove that the entries of the matrix A^{-1} depend smoothly (in fact analytically) on the entries of A.

*35. Give a proof that the inversion operator Inv is analytic (i.e., is defined locally by a convergent power series) as follows:

(a) If $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ and $\|T\| < 1$ show that the series of linear transformations

$$I + T + T^2 + \cdots + T^k + \ldots$$

converges to a linear transformation S, and

$$S \circ (I - T) = I = (I - T) \circ S,$$

where I is the identity transformation.

(b) Infer from (a) that inversion is analytic at I.

(c) In general, if $T_0 \in G$ and $\|T\| < 1/\left\|T_0^{-1}\right\|$, show that

$$\text{Inv}(T_0 - T) = \text{Inv}(I - T_0^{-1} \circ T) \circ T_0^{-1},$$

and infer that Inv is analytic at T_0.

(d) Infer from the general fact that analyticity implies smoothness that inversion is smooth.

(Note that this proof avoids Cramer's Rule and makes no use of finite dimensionality.)

*36. Give a proof of smoothness of Inv by the following bootstrap method.

(a) Using the identity

$$X^{-1} - Y^{-1} = X^{-1} \circ (Y - X) \circ Y^{-1}$$

give a simple proof that Inv is continuous.

(b) Infer that $Y = \text{Inv}(X)$ is a continuous solution of the C^∞ implicit function problem

$$F(X, Y) - I = 0,$$

where $F(X, Y) = X \circ Y$ as in Exercise 33. Since the proof of the C^1 Implicit Function Theorem relies only continuity of Inv, it is not circular reasoning to conclude that Inv is C^1.

(c) Assume simultaneously that the C^r Implicit Function Theorem has been proved, and that Inv is known to be C^{r-1}. Prove that Inv is C^r and that the C^{r+1} Implicit Function Theorem is true.

(d) Conclude logically that Inv is smooth and the C^∞ Implicit Function Theorem is true.

Note that this proof avoids Cramer's Rule and makes no use of finite dimensionality.

37. Draw pictures of all the possible shapes of $T(S^2)$ where $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear transformation and S^2 is the 2-sphere. (Don't forget the cases in which T has rank < 3.)

*38. Use polar decomposition to give an alternate proof of the volume multiplier formula.

**39. Consider the set S of all 2×2 matrices $X \in \mathcal{M}$ that have rank 1.

(a) Show that in a neighborhood of the matrix

$$X_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

S is diffeomorphic to a two-dimensional disc.

(b) Is this true (locally) for all matrices $X \in S$?

(c) Describe S globally. (How many connected components does it have? Is it closed? If not, what are its limit points and how does S approach them? What is the intersection of S with the unit sphere in $\mathcal{M}$?, etc.)

40. Let $0 \leq \epsilon \leq 1$ and $a, b > 0$ be given.

(a) Prove that

$$\left(\frac{1}{1+\epsilon}\right)^2 \leq \frac{a}{b} \leq (1+\epsilon)^2 \quad \Rightarrow \quad |a - b| \leq 16\epsilon b.$$

(b) Is the estimate in (a) sharp? (That is, can 16 be replaced by a smaller constant?)

41. Suppose that f and g are r^{th}-order differentiable and that the composite $h = g \circ f$ makes sense. A **partition divides a set into nonempty disjoint subsets. Prove the Higher-Order Chain Rule,

$$(D^r h)_p = \sum_{k=1}^{r} \sum_{\mu \in P(k,r)} (D^k g)_q \circ (D^\mu f)_p$$

where μ partitions $\{1, \ldots, r\}$ into k subsets, and $q = f(p)$. In terms of r-linear transformations, this notation means

$$(D^r h)_p(v_1, \ldots, v_r) = \sum_{k=1}^{r} \sum_{\mu} (D^k g)_q((D^{|\mu_1|} f)_p(v_{\mu_1}), \ldots, (D^{|\mu_k|} f)_p(v_{\mu_k}))$$

where $|\mu_i| = \#\mu_i$ and v_{μ_i} is the $|\mu_i|$-tuple of vectors v_j with $j \in \mu_i$. (Symmetry implies that the order of the vectors v_j in the $|\mu_i|$-tuple v_{μ_i} and the order in which the partition blocks $\mu_1, \ldots, \mu_k$ occur are irrelevant.)

**42. Suppose that β is bilinear and $\beta(f, g)$ makes sense. If f and g are r^{th} order differentiable at p, find the Higher-Order Leibniz Formula for $D^r(\beta(f, g))_p$. [Hint: First derive the formula in dimension 1.]

43. Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ has rank k.
   (a) Show that there exists a $\delta > 0$ such that if $S : \mathbb{R}^n \to \mathbb{R}^m$ and $\|S - T\| < \delta$, then S has rank $\geq k$.
   (b) Give a specific example in which the rank of S can be greater than the rank of T, no matter how small δ is.
   (c) Give examples of linear transformations of rank k for each k where $0 \leq k \leq \min\{n, m\}$.

44. Let $S \subset M$.
   (a) Define the characteristic function $\chi_S : M \to \mathbb{R}$.
   (b) If M is a metric space, show that $\chi_S(x)$ is discontinuous at x if and only if x is a boundary point of S.

45. On page 302 there is a definition of $Z \subset \mathbb{R}^2$ being a zero set that involves open rectangles.
   (a) Show that the definition is unaffected if we require that the rectangles covering Z are squares.
   (b) What if we permit the squares or rectangles to be non-open?
   (c) What if we use discs or other shapes instead of squares and rectangles?

*46. Assume that $S \subset \mathbb{R}^2$ is bounded.
   (a) Prove that if S is Riemann measurable then so are its interior and closure.
   (b) Suppose that the interior and closure of S are Riemann measurable and $|\text{int}(S)| = |\overline{S}|$. Prove that S is Riemann measurable.
   (c) Show that some open bounded subsets of $\mathbb{R}^2$ are not Riemann measurable.

*47. In the derivation of Fubini's Theorem on page 304, it is observed that for all $y \in [c, d] \setminus Y$, where Y is a zero set, the lower and upper integrals with respect to x agree, $\underline{F}(y) = \overline{F}(y)$. One might think that the values of $\underline{F}$ and $\overline{F}$ on Y have no effect on their integrals. Not so. Consider the function defined on the unit square $[0, 1] \times [0, 1]$,

$$f(x, y) = \begin{cases} 1 & \text{if } y \text{ is irrational} \\ 1 & \text{if } y \text{ is rational and } x \text{ is irrational} \\ 1 - \dfrac{1}{q} & \text{if } y \text{ is rational and } x = p/q \text{ is rational} \\ & \text{and written in lowest terms.} \end{cases}$$

   (a) Show that f is Riemann integrable, and its integral is 1.
   (b) Observe that if Y is the zero set $\mathbb{Q} \cap [0, 1]$, then for each $y \notin Y$,

$$\int_0^1 f(x, y)\, dx$$

   exists and equals 1.
   (c) Observe that if for each $y \in Y$ we choose in a completely arbitrary manner some

$$h(y) \in [\underline{F}(y), \overline{F}(y)]$$

   and set

$$H(x) = \begin{cases} \underline{F}(y) = \overline{F}(y) & \text{if } y \notin Y \\ h(y) & \text{if } y \in Y \end{cases}$$

   then the integral of H exists and equals 1, but if we take $g(x) = 0$ for all $y \in Y$ then the integral of

$$G(x) = \begin{cases} \underline{F}(y) = \overline{F}(y) & \text{if } y \notin Y \\ g(y) = 0 & \text{if } y \in Y \end{cases}$$

   does not exist.

***48. Is there a criterion to decide which redefinitions of the Riemann integral on the zero set Y of Exercise 47 are harmless and which are not?

49. Using the Fundamental Theorem of Calculus, give a direct proof of Green's Formulas

$$-\iint_R f_y\, dxdy = \int_{\partial R} f\, dx \quad \text{and} \quad \iint_R g_x\, dxdy = \int_{\partial R} g\, dy.$$

R is a square in the plane and $f, g : \mathbb{R}^2 \to \mathbb{R}$ are smooth. (Assume that the edges of the square are parallel to the coordinate axes.)

50. Draw a **staircase** curve S_n that approximates the diagonal $\Delta = \{(x, y) \in \mathbb{R}^2 : 0 \le x = y \le 1\}$ to within a tolerance $1/n$. (S_n consists of both treads and risers.) Suppose that $f, g : \mathbb{R}^2 \to \mathbb{R}$ are smooth.
    (a) Why does length $S_n \not\to$ length Δ?
    (b) Despite (a), prove $\int_{S_n} f\,dx \to \int_\Delta f\,dx$ and $\int_{S_n} g\,dy \to \int_\Delta g\,dy$.
    (c) Repeat (b) with Δ replaced by the graph of a smooth function $g : [a, b] \to \mathbb{R}$.
    (d) If C is a smooth simple closed curve in the plane, show that it is the union of finitely many arcs C_ℓ, each of which is the graph of a smooth function $y = g(x)$ or $x = g(y)$, and the arcs C_ℓ meet only at common endpoints.
    (e) Infer that if (S_n) is a sequence of staircase curves that converges to C then $\int_{S_n} f dx + g dy \to \int_C f dx + g dy$.
    (f) Use (e) and Exercise 49 to give a proof of Green's Formulas on a general region $D \subset \mathbb{R}^2$ bounded by a smooth simple closed curve C, that relies on approximating[†] C, say from the inside, by staircase curves S_n which bound regions R_n composed of many small squares. (You may imagine that $R_1 \subset R_2 \subset \ldots$ and that $R_n \to D$.)

51. A region R in the plane is of **type 1** if there are smooth functions $g_1 : [a, b] \to \mathbb{R}$, $g_2 : [a, b] \to \mathbb{R}$ such that $g_1(x) \le g_2(x)$ and

$$R = \{(x, y) : a \le x \le b \text{ and } g_1(x) \le y \le g_2(x)\}.$$

R is of **type 2** if the roles of x and y can be reversed, and it is **simple** if it is of both type 1 and type 2.
    (a) Give an example of a region that is type 1 but not type 2.
    (b) Give an example of a region that is neither type 1 nor type 2.
    (c) Is every simple region starlike? convex?
    (d) If a convex region is bounded by a smooth simple closed curve, is it simple?
    (e) Give an example of a region that divides into three simple subregions but not into two.

[†] This staircase approximation proof generalizes to regions that are bounded by fractal, nondifferentiable curves such as the von Koch snowflake. As Jenny Harrison has shown, it also generalizes to higher dimensions.

    *(f) If a region is bounded by a smooth simple closed curve C, it need not divide into a finite number of simple subregions. Find an example.
    (g) Infer that the standard proof of Green's Formulas for simple regions (as, for example, in J. Stewart's *Calculus*) does not immediately carry over to the general planar region R with smooth boundary, i.e., cutting R into simple regions can fail.
    *(h) Show that if the curve C in (f) is analytic, then no such example exists. [Hint: C is analytic if it is locally the graph of a function defined by a convergent power series. A nonconstant analytic function has the property that for each x, there is some derivative of f which is nonzero, $f^{(r)}(x) \neq 0$.]

**52. The 2-cell $\varphi : I^n \to B^n$ constructed in Step 3 of the proof of Brouwer's Theorem is smooth but not one-to-one.
    (a) Construct a homeomorphism $h : I^2 \to B^2$ where I^2 is the closed unit square and B^2 is the closed unit disc.
    (b) In addition make h in (a) be of class C^1 (on the closed square) and be a diffeomorphism from the interior of I^2 onto the interior of B^2. (The derivative of a diffeomorphism is everywhere nonsingular.)
    (c) Why can h not be a diffeomorphism from I^2 onto B^2?
    (d) Improve class C^1 in (b) to class C^∞.

53. If $K, L \subset \mathbb{R}^n$ and if there is a homeomorphism $h : K \to L$ that extends to $H : U \to V$ such that $U, V \subset \mathbb{R}^n$ are open, H is a homeomorphism, and H, H^{-1} are of class C^r, $1 \le r \le \infty$, then we say that K and L are **ambiently C^r-diffeomorphic.
    (a) In the plane, prove that the closed unit square is ambiently diffeomorphic to a general rectangle; to a general parallelogram.
    (b) If K, L are ambiently diffeomorphic polygons in the plane, prove that K and L have the same number of vertices. (Do not count vertices at which the interior angle is 180 degrees.)
    (c) Prove that the closed square and closed disc are not ambiently diffeomorphic.
    (d) If K is a convex polygon that is ambiently diffeomorphic to a polygon L, prove that L is convex.
    (e) Is the converse to (b) true or false? What about in the convex case?

(f) The closed disc is tiled by five ambiently diffeomorphic copies of the unit square as shown in Figure 121. Prove that it cannot be tiled by fewer.

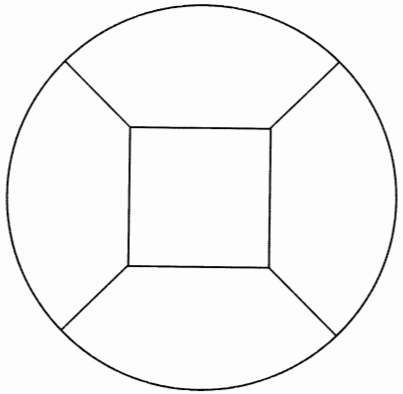


Figure 121 Five diffeomorphs of the square tile the disc.

(g) Generalize to dimension $n \geq 3$ and show that the n-ball can be tiled by $2n+1$ diffeomorphs of the n-cube. Can it be done with fewer?

(h) Show that a triangle can be tiled by three diffeomorphs of the square. Infer that any surface that can be tiled by diffeomorphs of the triangle can also be tiled by diffeomorphs of the square. What happens in higher dimensions?

54. Choose at random I, J, two triples of integers between 1 and 9. Check that $dx_I \wedge dx_J = dx_{IJ}$.
55. Show that $d : \Omega^k \to \Omega^{k+1}$ is a linear vector space homomorphism.
56. Prove that the pullback acts linearly on forms, and that it is natural with respect to composition in the sense that $(T \circ S)^* = S^* \circ T^*$. See how succinct a proof you can give.
57. True or false: if ω is a k-form and k is odd, then $\omega \wedge \omega = 0$. What if k is even and ≥ 2?
58. Does there exist a continuous mapping from the circle to itself that has no fixed point? What about the 2-torus? The 2-sphere?
59. Show that a smooth map $T : U \to V$ induces a linear map of cohomology groups $H^k(V) \to H^k(U)$ defined by

$$T^* : [\omega] \mapsto [T^*\omega].$$

Here, $[\omega]$ denotes the equivalence class of $\omega \in Z^k(V)$ in $H^k(V)$. The question amounts to showing that the pullback of a closed form ω is closed and that its cohomology class depends only on the cohomology class of ω.[†]

60. Prove that diffeomorphic open sets have isomorphic cohomology groups.
61. Show that the 1-form defined on $\mathbb{R}^2 \setminus \{(0,0)\}$ by

$$\omega = \frac{-y}{r^2}\,dx + \frac{x}{r^2}\,dy$$

is closed but not exact. Why do you think that this 1-form is often referred to as $d\theta$ and why is the name problematic?

62. Show that the 2-form defined on the spherical shell by

$$\omega = \frac{x}{r^3}\,dy \wedge dz + \frac{y}{r^3}\,dz \wedge dx + \frac{z}{r^3}\,dx \wedge dy$$

is closed but not exact.

63. If ω is closed, is the same true of $f\omega$? What if ω is exact?
64. Is the wedge product of closed forms closed? Of exact forms exact? What about the product of a closed form and an exact form? Does this give a ring structure to the cohomology classes?
65. Prove that the n-cell $\psi : [-1,1]^n \to B^n$ in the proof of the Brouwer Fixed Point Theorem has Jacobian $\rho'(r)\rho(r)^{n-1}/r^{n-1}$ for $r = |v|$ as claimed on page 336.

66. The **Hairy Ball Theorem states that any continuous vector field X in $\mathbb{R}^3$ that is everywhere tangent to the 2-sphere S is zero at some point of S. Here is an outline of a proof for you to fill in. (If you imagine the vector field as hair combed on a sphere, there must be a cowlick somewhere.)

(a) Show that the Hairy Ball Theorem is equivalent to a fixed point assertion: every continuous map of S to itself that is sufficiently close to the identity map $S \to S$ has a fixed point. (This is not needed below, but it is interesting.)

(b) If a continuous vector field on S has no zero on or inside a small simple closed curve $C \subset S$, show that the net angular turning of X along C as judged by an observer who takes a tour of C in the

[†] A fancier way to present the proof of the Brouwer Fixed Point Theorem goes like this: As always, the question reduces to showing that there is no smooth retraction T of the n-ball to its boundary. Such a T would give a cohomology map $T^* : H^k(\partial B) \to H^k(B)$ where the cohomology groups of ∂B are those of its spherical shell neighborhood. The map T^* is seen to be a cohomology group isomorphism because $T \circ \text{inclusion}_{\partial B} = \text{inclusion}_{\partial B}$ and $\text{inclusion}_{\partial B}{}^* = \text{identity}$. But when $k = n-1 \geq 1$ the cohomology groups are non-isomorphic; they are computed to be $H^{n-1}(\partial B) = \mathbb{R}$ and $H^{n-1}(B) = 0$.

counterclockwise direction is -2π. (The observer walks along C in the counterclockwise direction when S is viewed from the outside, and he measures the angle that X makes with respect to his own tangent vector as he walks along C. By convention, clockwise angular variation is negative.) Show also that the net turning is $+2\pi$ if the observer walks along C in the clockwise direction.

(c) If C_t is a continuous family of simple closed curves on S, $a \leq t \leq b$, and if X never equals zero at points of C_t, show that the net angular turning of X along C_t is independent of t. (This is a case of a previous exercise stating that a continuous integer valued function of t is constant.)

(d) Imagine the following continuous family of simple closed curves C_t. For $t = 0$, C_0 is the arctic circle. For $0 \leq t \leq 1/2$, the latitude of C_t decreases while its circumference increases as it oozes downward, becomes the equator, and then grows smaller until it becomes the antarctic circle when $t = 1/2$. For $1/2 \leq t \leq 1$, C_t maintains its size and shape, but its new center, the South Pole, slides up the Greenwich Meridian until at $t = 1$, C_t regains its original arctic position. See Figure 122. Its orientation has reversed. Orient the arctic circle C_0 positively and choose an orientation on each C_t that depends continuously on t. To reach a contradiction, suppose that X has no zero on S.

(i) Why is the total angular turning of X along C_0 equal to -2π?.

(ii) Why is it $+2\pi$ on C_1?

(iii) Why is this a contradiction to (c) unless X has a zero somewhere?

(iv) Conclude that you have proved the Hairy Ball Theorem.

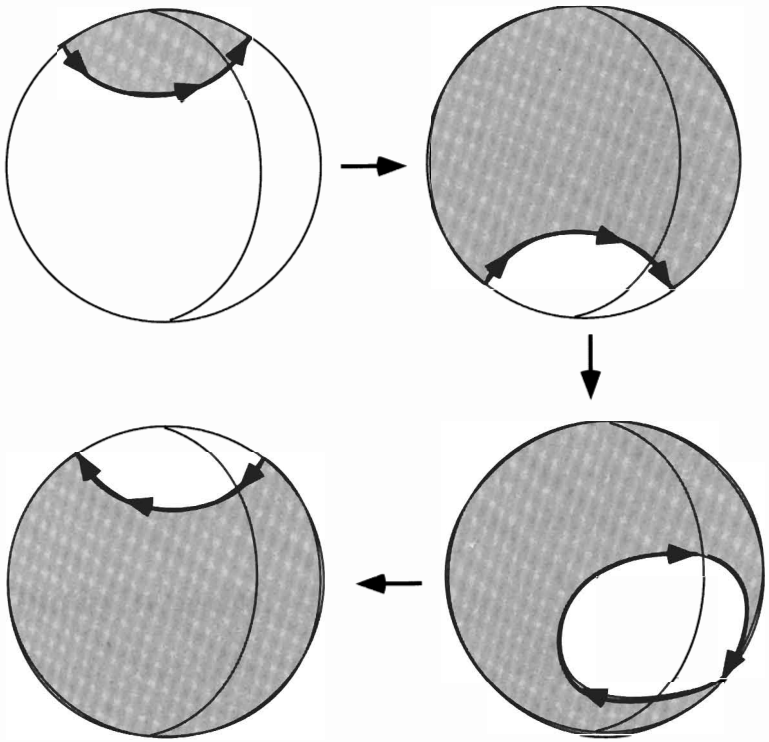


Figure 122 A deformation of the arctic circle that reverses its orientation.

6

Lebesgue Theory

This chapter presents a geometric theory of Lebesgue measure and integration. In calculus you certainly learned that the integral is the area under the curve. With a good definition of area, that is the point of view we advance here. Deriving the basic theory of Lebesgue integration then becomes a matter of inspecting the right picture.

1 Outer measure

How should you measure the length of a subset of the line? If the set to be measured is simple, so is the answer: the length of the interval (a, b) is $b-a$. But what is the length of the set of rational numbers? of the Cantor set? As is often the case in analysis we proceed by inequalities and limits. In fact one might distinguish the fields of algebra and analysis solely according to their use of equalities versus inequalities.

Definition The **length** of an interval $I = (a, b)$ is $b - a$. It is denoted $|I|$. The **Lebesgue outer measure** of a set $A \subset \mathbb{R}$ is

$$m^* A = \inf\{\sum_k |I_k| : \{I_k\} \text{ is a covering of } A \text{ by open intervals}\}.$$

Tacitly we assume that the covering is countable; the series $\sum_k |I_k|$ is its **total length**. (Recall that "countable" means either finite or denumerable.)

The outer measure of A is the infimum of the total lengths of all possible coverings $\{I_k\}$ of A by open intervals. If every series $\sum_k |I_k|$ diverges, then by definition, $m^*A = \infty$.

Outer measure is defined for every $A \subset \mathbb{R}$. It measures A from the outside as do calipers. A dual approach measures A from the inside. It is called **inner measure**, is denoted m_*A, and is discussed in Section 3.

Three properties of outer measure (the **axioms of outer measure**) are easy to check.

(a) The outer measure of the empty set is 0, $m^*\emptyset = 0$.

(b) If $A \subset B$ then $m^*A \le m^*B$.

(c) If $A = \bigcup_{n=1}^{\infty} A_n$ then $m^*A \le \sum_{n=1}^{\infty} m^*A_n$.

(b) and (c) are called **monotonicity** and **countable subadditivity** of outer measure.

(a) is obvious. Every interval covers the empty set.

(b) is obvious. Every covering of B is also a covering of A.

(c) uses the $\epsilon/2^n$ trick. Given $\epsilon > 0$ there exists for each n a covering $\{I_{kn} : k \in \mathbb{N}\}$ of A_n such that

$$\sum_{k=1}^{\infty} |I_{kn}| < m^*A_n + \frac{\epsilon}{2^n}.$$

The collection $\{I_{kn} : k, n \in \mathbb{N}\}$ covers A and

$$\sum_{k,n} |I_{kn}| = \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} |I_{kn}| \le \sum_{n=1}^{\infty} \left(m^*A_n + \frac{\epsilon}{2^n}\right) = \sum_{n=1}^{\infty} m^*A_n + \epsilon.$$

Thus the infimum of the total lengths of coverings of A by open intervals is $\le \sum_n m^*A_n + \epsilon$, and since $\epsilon > 0$ is arbitrary, the infimum is $\le \sum_n m^*A_n$, which is what (c) asserts.

Next, suppose you have a set A in the plane and you want to measure its area. Here is the natural way to do it.

Definition The **area** of a rectangle $R = (a, b) \times (c, d)$ is $|R| = (b - a) \cdot (d - c)$, and the **(planar) outer measure** of $A \subset \mathbb{R}^2$ is the infimum of the total area of countable coverings of A by open rectangles R_k

$$m^*A = \inf\{\sum_k |R_k| : \{R_k\} \text{ covers } A\}.$$

If need be, we decorate $|\ |$ and m^* with subscripts "1" and "2" to distinguish the linear and planar quantities.

The outer measure axioms — monotonicity, countable subadditivity, and the outer measure of the empty set being zero — are true for planar outer measure too. See also Exercise 5.

A consequence of subadditivity concerns **zero sets**, sets that have zero outer measure.

1 Proposition *The countable union of zero sets is a zero set.*

Proof If $m^*Z_n = 0$ for all $n \in \mathbb{N}$ and $Z = \bigcup Z_n$ then by (c),

$$m^*Z \le \sum_n m^*Z_n = 0.$$

Two other properties enjoyed by outer measure are left for you as Exercises 1 and 2:

(d) Rigid translation of a set has no effect on its outer measure.

(e) Dilation of a set dilates its outer measure correspondingly.

The next theorem states a property of outer measure that seems obvious. See also Exercises 3, 4.

2 Theorem *The linear outer measure of a closed interval is its length; the planar outer measure of a closed rectangle is its area.*

Proof Let $I = [a, b]$ be the interval. For any $\epsilon > 0$, the single open interval $(a - \epsilon, b + \epsilon)$ covers I, so $m^*I \le |I| + 2\epsilon$. Since ϵ is arbitrary,

$$m^*I \le |I|.$$

To check the reverse inequality, let $\{I_k\}$ be a countable covering of I by open intervals. Compactness implies that for some finite N, $I \subset \bigcup_{k=1}^{N} I_k$. Form a partition

$$a = x_0 < \cdots < x_n = b$$

such that all the endpoints a_k, b_k of the intervals $I_k = (a_k, b_k)$ that appear in I occur as partition points. Each partition interval $X_i = (x_{i-1}, x_i)$ is contained wholly inside some covering interval I_k, which implies

$$|X_i| \le \sum_{k=1}^{N} |X_i \cap I_k|.$$

For one of the terms in the sum is $|X_i|$ itself. Also, for each k, the partition of I restricts to a partition of the interval $I \cap I_k$, and thus

$$|I \cap I_k| = \sum_{i=1}^{n} |X_i \cap I_k|.$$

This gives

$$|I| = \sum_{i=1}^{n} |X_i| \le \sum_{i=1}^{n} \Big(\sum_{k=1}^{N} |X_i \cap I_k| \Big) = \sum_{k=1}^{N} \Big(\sum_{i=1}^{n} |X_i \cap I_k| \Big)$$
$$= \sum_{k=1}^{N} |I \cap I_k| \le \sum_{k=1}^{N} |I_k| \le \sum_{k=1}^{\infty} |I_k| ,$$

which shows that $|I| \le m^* I$ and completes the proof in dimension one.

The case of a rectangle $R = [a, b] \times [c, d]$ is similar. As with intervals, it is enough to show that any countable covering of R by open rectangles $R_k = (a_k, b_k) \times (c_k, d_k)$ has $|R| \le \sum |R_k|$. There is an N such that $R_1, \ldots, R_N$ cover R. Partition the sides of R as

$$a = x_0 < \cdots < x_n = b \quad c = y_0 < \cdots < y_m = d$$

so that all the endpoints a_k, b_k, c_k, d_k of the covering intervals that appear in $[a, b]$ and $[c, d]$ occur among the partition points. Call $X_i = (x_{i-1}, x_i)$, $Y_j = (y_{j-1}, y_j)$, and $G_{ij} = X_i \times Y_j$. Each of these grid rectangles G_{ij} is contained wholly inside at least one covering rectangle R_k, and so $|G_{ij}| \le \sum_{k=1}^{N} |G_{ij} \cap R_k|$. Also, for each k the rectangle $R \cap R_k$ is partitioned by the grid rectangles that lie in it so that $|R \cap R_k| = \sum_{ij} |G_{ij} \cap (R \cap R_k)|$. This gives

$$|R| = \Big(\sum_{i=1}^{n} |X_i| \Big) \Big(\sum_{j=1}^{m} |Y_j| \Big) = \sum_{ij} |G_{ij}| \le \sum_{ij} \Big(\sum_{k=1}^{N} |G_{ij} \cap R_k| \Big)$$
$$= \sum_{k=1}^{N} \Big(\sum_{ij} |G_{ij} \cap (R \cap R_k)| \Big) = \sum_{k=1}^{N} |R \cap R_k| \le \sum_{k=1}^{N} |R_k| \le \sum_{k=1}^{\infty} |R_k| ,$$

which shows that $|R| \le m^* R$ and completes the proof. □

3 Corollary *The formulas $m^* I = |I|$ and $m^* R = |R|$ hold also for intervals and rectangles that are open or partly open.*

Proof If I is any interval and $\epsilon > 0$ is given then there are closed intervals J, J' that sandwich I as $J \subset I \subset J'$ and $|J'| - |J| < \epsilon$. Then

$$\begin{array}{ccccc} m^* J & \le & m^* I & \le & m^* J' \\ \| & & & & \| \\ |J| & \le & |I| & \le & |J'| . \end{array}$$

Since $\big||I| - m^* I\big| < \epsilon$ for all $\epsilon > 0$, $m^* I = |I|$. The sandwich method works equally well for rectangles. □

2 Measurability

If A and B are subsets of disjoint intervals in $\mathbb{R}$ it is easy to show that

$$m^*(A \sqcup B) = m^* A + m^* B.$$

But what if A and B are merely disjoint? Is the formula still true? The answer is "yes," if the sets have an additional property called measurability, and "no" in general as is shown in Appendix B. Measurability is the rule and nonmeasurability the exception: the sets you meet in analysis — open sets, closed sets, their unions, differences, etc. — all are measurable. See Section 3.

Definition A set $E \subset \mathbb{R}$ is **(Lebesgue) measurable** if the division $E|E^c$ of $\mathbb{R}$ is so "clean" that for each "test set" $X \subset \mathbb{R}$,

$$m^* X = m^*(X \cap E) + m^*(X \cap E^c). \tag{1}$$

We denote by $\mathcal{M} = \mathcal{M}(\mathbb{R})$ the collection of all Lebesgue measurable subsets of $\mathbb{R}$. If E is measurable its **Lebesgue measure** is $m^* E$, which we write as mE, dropping the asterisk to emphasize measurability of E.

The definition of measurability in the plane is analogous: if $E \subset \mathbb{R}^2$ is measurable, then $E|E^c$ divides each $X \subset \mathbb{R}^2$ so cleanly that (1) is true for the planar outer measures.

Which sets are measurable? It is obvious that the empty set is measurable. It is also obvious that if a set is measurable, then so is its complement, since $E|E^c$ and $E^c|E$ divide a test set X in the same way. In this section we analyze measurability in the abstract. For the basic facts about measurability have nothing to do with $\mathbb{R}$ or $\mathbb{R}^2$. They hold for any "abstract outer measure."

Definition Let M be any set. The collection of all subsets of M is denoted as 2^M. An **abstract outer measure** on M is a function $\omega : 2^M \to [0, \infty]$ that satisfies the three axioms of outer measure: $\omega(\emptyset) = 0$, ω is monotone, and ω is countably subadditive. A set $E \subset M$ is **measurable** with respect to ω if $M = E|E^c$ is so clean that for each test set $X \subset M$,

$$\omega X = \omega(X \cap E) + \omega(X \cap E^c).$$

4 Theorem *The collection $\mathcal{M}$ of measurable sets with respect to any outer measure on any set M is a σ-algebra and the outer measure restricted to this σ-algebra is countably additive. In particular, Lebesgue measure is countably additive.*

A **σ-algebra** is a collection of sets that includes the empty set, is closed under complement, and is closed under countable union. **Countable additivity** of ω means that if $E_1, E_2, \ldots$ are measurable with respect to ω, then

$$E = \bigsqcup_i E_i \quad \Rightarrow \quad \omega E = \sum_i \omega E_i.$$

Proof Let $\mathcal{M}$ denote the collection of measurable sets with respect to the outer measure ω on M. To check that $\mathcal{M}$ is a σ-algebra we must show that it contains the empty set, is closed under complements, and is closed under countable union.

It is clear that $\emptyset$ divides any X cleanly, so $\emptyset \in \mathcal{M}$. Also, since $E|E^c$ divides a test set X in the same way that $E^c|E$ does, $\mathcal{M}$ is closed under complements. To check that $\mathcal{M}$ is closed under countable union takes four preliminary steps:

(a) $\mathcal{M}$ is closed under differences.
(b) $\mathcal{M}$ is closed under finite union.
(c) ω is finitely additive on $\mathcal{M}$.
(d) ω satisfies a special countable addition formula.

(a) For measurable sets E_1, E_2, and a test set X, draw the Venn diagram in Figure 123 where X is represented as a disc. To check measurability of $E_1 \setminus E_2$ we must verify the equation

$$2 + 134 = 1234$$

where $2 = \omega(X \cap (E_1 \setminus E_2))$, $134 = \omega(X \cap (E_1 \setminus E_2))^c$, $1234 = \omega X$, etc. Since E_1 divides any set cleanly, $134 = 1 + 34$, and since E_2 divides any set cleanly, $34 = 3 + 4$. Thus,

$$2 + 134 = 2 + 1 + 3 + 4 = 1 + 2 + 3 + 4.$$

For the same reason, $1234 = 12 + 34 = 1 + 2 + 3 + 4$, which completes the proof of (a).

(b) Suppose that E_1, E_2 are measurable and $E = E_1 \cup E_2$. Since $E^c = E_1^c \setminus E_2$, (a) implies that $E^c \in \mathcal{M}$, and thus $E \in \mathcal{M}$. For more than two sets, induction shows that if $E_1, \ldots, E_n \in \mathcal{M}$ then $E_1 \cup \cdots \cup E_n \in \mathcal{M}$.

(c) If $E_1, E_2 \in \mathcal{M}$ are disjoint, then E_1 divides $E = E_1 \sqcup E_2$ cleanly, so

$$\omega E = \omega(E \cap E_1) + \omega(E \cap E_1^c) = \omega(E_1) + \omega(E_2),$$

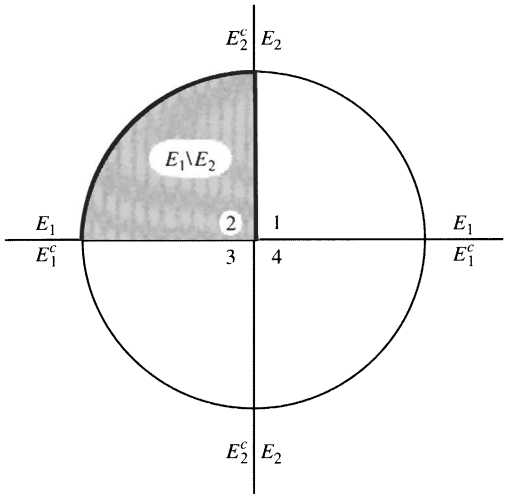


Figure 123 The picture that proves $\mathcal{M}$ is closed under differences.

which is additivity for pairs of measurable sets. For more than two measurable sets, induction implies that ω is **finitely additive** on $\mathcal{M}$; i.e., if $E_1, \ldots, E_n \in \mathcal{M}$ then

$$E = \bigsqcup_{i=1}^n E_i \quad \Rightarrow \quad \omega E = \sum_{i=1}^n \omega(E_i).$$

(d) Given a test set $X \subset M$, and a countable disjoint union of measurable sets $E = \bigsqcup E_i$ of measurable sets, we claim that

$$\omega(X \cap E) = \sum_i \omega(X \cap E_i). \tag{2}$$

(When $X = M$ this is countable additivity, but in general X need not be measurable.) Consider the division

$$X \cap (E_1 \sqcup E_2) = (X \cap E_1) \sqcup (X \cap E_2).$$

Measurability of E_1 implies that the two outer measures add. By induction the same is true for any finite sum,

$$\omega(X \cap (E_1 \sqcup \cdots \sqcup E_k)) = \omega(X \cap E_1) + \cdots + \omega(X \cap E_k).$$

Monotonicity of ω implies that

$$\omega(X \cap E) \geq \omega(X \cap (E_1 \sqcup \cdots \sqcup E_k)),$$

and so $\omega(X \cap E)$ dominates each partial sum of the series $\sum \omega(X \cap E_i)$. Hence it dominates the series too,

$$\sum_{i=1}^{\infty} \omega(X \cap E_i) \leq \omega(X \cap E).$$

The reverse inequality is always true by subadditivity and we get equality, verifying (2).

Finally, we prove that $E = \bigcup E_i$ is measurable when each E_i is. Taking $E_i' = E_i \setminus (E_1 \cup \cdots \cup E_{i-1})$, (a) tells us that it is no loss of generality to assume the sets E_i are disjoint, $E = \bigsqcup E_i$. Given a test set $X \subset M$ we know by (c) (finite additivity) and monotonicity of ω that

$$\begin{aligned} & \omega(X \cap E_1) + \cdots + \omega(X \cap E_k) + \omega(X \cap E^c) \\ = \; & \omega(X \cap (E_1 \sqcup \cdots \sqcup E_k)) + \omega(X \cap E^c) \\ \leq \; & \omega(X \cap (E_1 \sqcup \cdots \sqcup E_k)) + \omega(X \cap (E_1 \sqcup \cdots \sqcup E_k)^c) \\ = \; & \omega X. \end{aligned}$$

Being true for all k, the inequality holds also for the full series,

$$\sum_{i=1}^{\infty} \omega(X \cap E_i) + \omega(X \cap E^c) \leq \omega X.$$

Using (2), we get

$$\omega(X \cap E) + \omega(X \cap E^c) = \sum_{i=1}^{\infty} \omega(X \cap E_i) + \omega(X \cap E^c) \leq \omega X.$$

The reverse inequality is true by subadditivity of ω. This gives equality and shows that E is measurable. Hence $\mathcal{M}$ is a σ-algebra and the restriction of ω to $\mathcal{M}$ is countably additive. □

From countable additivity we deduce a very useful fact about measures. It applies to any outer measure ω, in particular to Lebesgue outer measure.

5 Measure Continuity Theorem *If $\{E_k\}$ and $\{F_k\}$ are sequences of measurable sets then*

upward measure continuity	$E_k \uparrow E \Rightarrow \omega E_k \uparrow \omega E$
downward measure continuity	$F_k \downarrow F$ *and* $\omega F_1 < \infty \Rightarrow \omega F_k \downarrow \omega F$.

Proof The notation $E_k \uparrow E$ means that $E_1 \subset E_2 \subset \ldots$ and $E = \bigcup E_k$. Write E disjointly as $E = \bigsqcup E_k'$ where $E_k' = E_k \setminus (E_1 \cup \cdots \cup E_{k-1})$. Countable additivity gives

$$\omega E = \sum_{n=1}^{\infty} \omega E_n'.$$

Also, the k^{th} partial sum of the series equals ωE_k, so ωE_k converges upward to ωE. The notation $F_k \downarrow F$ means that $F_1 \supset F_2 \supset \ldots$ and $F = \bigcap F_k$. Write F_1 disjointly as

$$F_1 = \Big(\bigsqcup_{k=1}^{\infty} F_k'\Big) \sqcup F$$

where $F_k' = F_k \setminus F_{k+1}$. Then $F_k = \bigsqcup_{n \geq k} F_n' \sqcup F$. The countable additivity formula

$$\omega F_1 = \omega F + \sum_{n=1}^{\infty} \omega F_n'$$

plus finiteness of ωF_1 implies that the series converges to a finite limit, so its tails converge to zero. That is

$$\omega F_k = \sum_{n=k}^{\infty} \omega F_n' + \omega F$$

converges downward to ωF as $k \to \infty$. □

3 Regularity

In this section we discuss properties of Lebesgue measure related to the topology of $\mathbb{R}$ and $\mathbb{R}^2$.

6 Theorem *Open sets and closed sets are measurable.*

7 Proposition *The inclusion or exclusion of zero sets has no effect on outer measure or measurability.*

Proof If Z is a zero set then we assert that $m^*(E \cup Z) = m^*E = m^*(E \setminus Z)$ and further, that if E is measurable then so are $E \cup Z$ and $E \setminus Z$. We have

$$m^*E \leq m^*(E \cup Z) \leq m^*E + m^*Z = m^*E,$$

so $m^*(E \cup Z) = m^*E$. Applying this to the set $E \setminus Z$ gives $m^*(E \setminus Z) = m^*((E \setminus Z) \cup (E \cap Z))$; i.e, $m^*E = m^*(E \setminus Z)$.

Assume that E is measurable. Given a test set X we use subadditivity, monotonicity, and the set theory formula $(E \cup Z)^c = E^c \setminus Z$ to get

$$\begin{aligned} m^*X &\le m^*(X \cap (E \cup Z)) + m^*(X \cap (E^c \setminus Z)) \\ &\le m^*(X \cap E) + m^*(X \cap Z) + m^*(X \cap E^c) = m^*X. \end{aligned}$$

Thus the inequalities are equalities and $E \cup Z$ is measurable. Applying this to E^c shows that $E^c \cup Z = (E \setminus Z)^c$ is measurable, so $E \setminus Z$ is measurable. □

8 Corollary *Zero sets are measurable.*

Proof Take $Z = \emptyset \cup Z$ in Proposition 7. □

9 Proposition *The half line $H = [0, \infty) \subset \mathbb{R}$ and the half plane $H \times \mathbb{R} \subset \mathbb{R}^2$ are measurable.*

Proof Given $X \subset \mathbb{R}$, we claim that $m^*X = m^*(X \cap H) + m^*(X \cap H^c)$. By Proposition 7 we may assume that $0 \notin X$, since X and $X \setminus \{0\}$ differ by the zero set $\{0\}$. Then

$$X = X^+ \sqcup X^- = (X \cap H) \sqcup (X \cap H^c).$$

Given $\epsilon > 0$ there is a covering $\mathcal{I}$ of X by open intervals I_k whose total length is

$$\sum_k |I_k| \le m^*X + \epsilon.$$

Split any interval I_k that contains 0 into its positive and negative open subintervals $I_k^{\pm}$. This gives a new covering $\mathcal{J}$ of X with the same total length as $\mathcal{I}$. Each interval in $\mathcal{J}$ lies wholly in the positive or negative half axis, so $\mathcal{J}$ splits into disjoint coverings $\mathcal{J}^{\pm}$ of $X^{\pm}$. Thus

$$\begin{aligned} m^*X &\le m^*(X \cap H) + m^*(X \cap H^c) \\ &\le \sum_{J \in \mathcal{J}^+} |J| + \sum_{J \in \mathcal{J}^-} |J| = \sum_k |I_k| \le m^*X + \epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary this gives measurability of H.

The planar case is similar. The y-axis has planar outer measure zero since for any $\epsilon > 0$ it is covered by the countable set of open rectangles $(-\epsilon/4^n, \epsilon/4^n) \times (-2^n, 2^n)$ whose total area is 4ϵ. By Proposition 7 it is then fair to exclude the y-axis from a test set $X \subset \mathbb{R}^2$, and the rest of the reasoning is the same as in the linear case. □

Proof of Theorem 6 Consider $\mathbb{R}$. By Proposition 9 the half lines $[0, \infty)$ and $(-\infty, 0)$ are measurable. Measurability is unaffected by translation, so $[a, \infty)$ and $(-\infty, b)$ are measurable. Measurability is also unaffected by the exclusion of zero sets. Thus (a, ∞) is measurable and so is

$$(a, b) = (a, \infty) \cap (-\infty, b).$$

Consider $\mathbb{R}^2$. The translates of half planes are measurable and it is fair to exclude zero sets. Thus the vertical strip $(a, b) \times \mathbb{R}$ is measurable. Exchanging the roles of the x-and y-axis shows that the horizontal strip $\mathbb{R} \times (c, d)$ is also measurable, so their intersection $(a, b) \times (c, d)$ is measurable.

Every open set in $\mathbb{R}$ is the countable union of open intervals, and every open set in the plane is the countable union of open rectangles. Since $\mathcal{M}$ is a σ-algebra, every open set is measurable, and since the complement of a measurable set is measurable, every closed set is also measurable. □

10 Corollary *The Lebesgue measure of an interval is its length and the Lebesgue measure of a rectangle is its area.*

Proof By Theorem 6, the interval and rectangle are measurable, so their measures equal their outer measures, which we know to be their length and area. □

Sets that are slightly more general than open sets and closed sets arise naturally. A countable intersection of open sets is called a **G_δ-set** and a countable union of closed sets is an **F_σ-set**. ("δ" stands for the German word *durschnitt* and "σ" stands for "sum.") By deMorgan's Laws, the complement of a G_δ-set is an F_σ-set and conversely. Since the σ-algebra of measurable sets contains the open sets and the closed sets, it also contains the G_δ-sets and the F_σ-sets.

11 Theorem *Lebesgue measure is **regular** in the sense that each measurable set E can be sandwiched between an F_σ-set and a G_δ-set, $F \subset E \subset G$, such that $m(G \setminus F) = 0$.*

Proof Assume first that $E \subset \mathbb{R}$ is bounded. There is a large closed interval I that contains E. Measurability implies that

$$mI = mE + m(I \setminus E).$$

There are decreasing sequences of open sets U_n and V_n such that $U_n \supset E$, $V_n \supset (I \setminus E)$, $mU_n \to mE$, and $mV_n \to m(I \setminus E)$ as $n \to \infty$. The

complements $K_n = I \setminus V_n$ form an increasing sequence of closed subsets of E and

$$mK_n = mI - mV_n \to mI - m(I \setminus E) = mE.$$

Thus $F = \bigcup K_n$ is an F_σ-set in E with $mF = mE$. Similarly there is a G_δ-set G that contains E and has $mG = mE$. Because all the measures are finite, the equality $mF = mE = mG$ implies that $m(G \setminus F) = 0$. The same reasoning applies to bounded subsets of the plane or $\mathbb{R}^n$. The unbounded case is left as Exercise 9. □

12 Corollary *Modulo zero sets, Lebesgue measurable sets are F_σ-sets and/or G_δ-sets.*

Proof $E = F \cup Z = G \setminus Z'$ for the zero sets $Z = E \setminus F$ and $Z' = G \setminus E$. □

The outer measure of a set $A \subset \mathbb{R}$ or $A \subset \mathbb{R}^n$ is the infimum of the measure of open sets that contain it. Dually, we define the **inner measure** of A to be the supremum of the measure of closed sets contained in it. We denote the inner measure of A as $m_* A$. Clearly $m_* A \leq m^* A$ and m_* measures A from the inside. Also, m_* is monotone: $A \subset B$ implies $m_* A \leq m_* B$.

Equivalently, $m^* A$ is the measure of the smallest measurable set $H \supset A$ while $m_* A$ is the measure of the largest measurable set $N \subset A$. (H being "smallest" means that for any other measurable $H' \supset A$, $H \setminus H'$ is a zero set. Similarly, if N' is any measurable subset of A then $N' \setminus N$ is a zero set. It is easy to see that H and N exist and are unique up to zero sets. They are called the **hull** and **kernel** of A. See Exercise 8.)

13 Theorem *A bounded set $A \subset \mathbb{R}$ (or $A \subset \mathbb{R}^n$) is Lebesgue measurable if and only if its inner and outer measures are equal, $m_* A = m^* A$. Further, if B is a bounded measurable set that contains A, then A is measurable if and only if it divides B cleanly.*

Proof If A is measurable then it is both its own largest measurable subset and the smallest measurable subset that contains it, so $m_* A = m^* A$.

On the other hand, if $m_* A = m^* A$, then A is sandwiched between measurable sets with equal finite measure, so it differs from each by a zero set and is therefore measurable.

Finally suppose that $B \supset A$ with B bounded and measurable. Then

$$m_* A + m^*(B \setminus A) = mB.$$

See Exercise 8. If A divides B cleanly, then

$$m^* A + m^*(B \setminus A) = mB.$$

Finiteness of all the measures implies that $m_* A = m^* A$ and A is measurable. The converse is obvious because a measurable set divides *every* test set cleanly. □

Remark Lebesgue took Theorem 13 as his definition of measurability. He said that a bounded set is measurable if its inner and outer measures are equal, and an unbounded set is measurable if it is a countable union of bounded measurable sets. In contrast, the current definition which uses cleanness and test sets is due to Caratheodory. It is easier to use (there are fewer complements to consider), unboundedness has no effect on it, and it generalizes more easily to "abstract measure spaces."

Regularity of Lebesgue measure has a number of uses. Here is one. See also Exercises 32 and 33.

14 Theorem *The Cartesian product of measurable sets is measurable, and the measure of the product is the product of the measures.*

Proof By convention, $0 \cdot \infty = 0 = \infty \cdot 0$. Given measurable $A, B \subset \mathbb{R}$, we claim that $A \times B$ is measurable with respect to planar measure and $m_2(A \times B) = m_1 A \cdot m_1 B$.

Case 1. A or B is a zero set; say it is A. Given $\epsilon > 0$ and an interval (c, d) there is a covering of A by open intervals $\{I_k\}$ with $\sum_k |I_k| < \epsilon/(d - c)$. Thus $A \times (c, d)$ is covered by rectangles $I_k \times (c, d)$ whose total area is $< \epsilon$, so $A \times (c, d)$ is a zero set. Since $A \times \mathbb{R} = \bigcup_{n=1}^{\infty} A \times (-n, n)$, it is a zero set, and since $A \times B \subset A \times \mathbb{R}$, $A \times B$ is also a zero set. Zero sets are measurable, which completes the proof of the theorem in this case.

Case 2. A and B are open sets. Then $A = \bigsqcup A_i$ and $B = \bigsqcup B_j$ where A_i and B_j are open intervals, and $A \times B$ is the disjoint countable union of open rectangles. It is therefore measurable and by countable additivity,

$$\begin{aligned} m(A \times B) &= \sum_{ij} |A_i \times B_j| = \sum_{ij} |A_i|\,|B_j| \\ &= \Big(\sum_i |A_i|\Big)\Big(\sum_j |B_j|\Big) = m_1 A \cdot m_1 B. \end{aligned}$$

Case 3. A and B are bounded G_δ-sets. Then there are nested decreasing sequences of open sets, $U_n \downarrow A$ and $V_n \downarrow B$ with $m_1 U_1 < \infty$ and

$m_1 V_1 < \infty$. Thus $U_n \times V_n \downarrow A \times B$ as $n \to \infty$. Being a G_δ-set, $A \times B$ is measurable. Downward measure continuity gives $m_1 U_n \to m_1 A, m_1 V_n \to m_1 B$, and by Case 2

$$\begin{array}{c} m_2(U_n \times V_n) = m_1 U_n \cdot m_1 V_n \to m_1 A \cdot m_1 B \\ \downarrow \\ m_2(A \times B) \end{array}$$

as $n \to \infty$.

Case 4. A and B are bounded measurable sets. By regularity, A and B differ from G_δ-sets G and H by zero sets X and Y. Thus $A \times B$ differs from $G \times H$ by the zero set

$$X \times H \quad \cup \quad G \times Y,$$

which implies that $A \times B$ is measurable and

$$m_2(A \times B) = m_2(G \times H) = m_1 G \cdot m_1 H = m_1 A \cdot m_1 B.$$

Case 5. A and B are measurable but not necessarily bounded. The proof of the theorem in this case is left to the reader. See Exercise 9. • □

See Exercise 13 for the n-dimensional version of Theorem 14.

4 Lebesgue integrals

Following J.C. Burkill, we justify the maxim that the integral of a function is the area under its graph. Let $f : \mathbb{R} \to [0, \infty)$ be given.

Definition The **undergraph** of f is

$$\mathcal{U}f = \{(x, y) \in \mathbb{R} \times [0, \infty) : 0 \le y < f(x)\}.$$

The function f is **Lebesgue measurable** if $\mathcal{U}f$ is measurable with respect to planar Lebesgue measure m_2, and if it is, then the **Lebesgue integral** of f is the measure of the undergraph

$$\int f = m_2(\mathcal{U}f).$$

When $X \subset \mathbb{R}$ and $f : X \to [0, \infty)$, the same definition applies: if the undergraph $\{(x, y) \in X \times [0, \infty) : 0 \le y < f(x)\}$ is measurable, then its measure is the Lebesgue integral of f over X,

$$\int_X f = m_2(\mathcal{U}f).$$

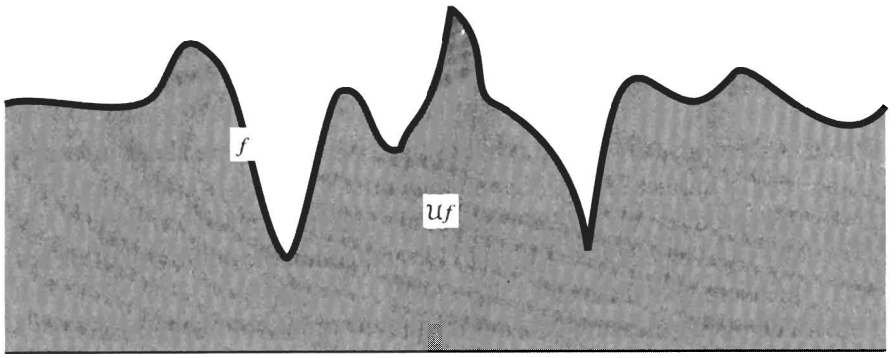


Figure 124 The geometric definition of the integral as the measure of the undergraph.

See Figure 124.

Burkill refers to the undergraph as the **ordinate set** of f. The notation for the Lebesgue integral intentionally omits the usual "dx" and the limits of integration to remind you that it is not merely the ordinary Riemann integral $\int_a^b f(x)\,dx$ or the improper Riemann integral $\int_{-\infty}^{\infty} f(x)\,dx$. The subscript "2" on the measure indicates planar measure and will usually be dropped.

Since a measurable set can have infinite measure, $\int f = \infty$ is permitted.

Definition The function $f : \mathbb{R} \to [0, \infty)$ is **Lebesgue integrable** if (it is measurable and) its integral is finite. The set of integrable functions is denoted by L^1, $\mathcal{L}^1$, or $\mathcal{L}$.

15 Lebesgue Monotone Convergence Theorem *Assume that $f_n : \mathbb{R} \to [0, \infty)$ is a sequence of measurable functions and $f_n \uparrow f$ as $n \to \infty$. Then*

$$\int f_n \uparrow \int f.$$

Proof Obvious. $\mathcal{U}f_n \uparrow \mathcal{U}f$ and measure continuity (Theorem 5) implies $m(\mathcal{U}f_n) \uparrow m(\mathcal{U}f)$. □

16 Theorem *Let $f, g : \mathbb{R} \to [0, \infty)$ be measurable functions.*

(a) If $f \le g$ then $\int f \le \int g$.

(b) If $\mathbb{R} = \bigsqcup_{k=1}^{\infty} X_k$ and each X_k is measurable, then

$$\int f = \sum_{k=1}^{\infty} \int_{X_k} f.$$

(c) If $X \subset \mathbb{R}$ is measurable, then $mX = \int \chi_X$.

(d) If $mX = 0$, then $\int_X f = 0$.

(e) If $f(x) = g(x)$ almost everywhere, then $\int f = \int g$.
(f) If $a \geq 0$, then $\int af = a \int f$.
(g) $\int f + g = \int f + \int g$.

Proof Assertions (a) - (f) are obvious from what we know about measure.
(a) $f \leq g$ implies $\mathcal{U}f \subset \mathcal{U}g$ and thus $m(\mathcal{U}f) \leq m(\mathcal{U}g)$.
(b) The product $X_k \times \mathbb{R}$ is measurable and its intersection with $\mathcal{U}f$ is $\mathcal{U}f|_{X_k}$. Thus, $\mathcal{U}f = \bigsqcup_{k=1}^{\infty} \mathcal{U}f|_{X_k}$ and countable additivity of planar measure give the result.
(c) The planar measure of the product $\mathcal{U}(\chi_X) = X \times [0, 1)$ is mX.
(d) $\mathcal{U}f$ is contained in the product $X \times \mathbb{R}$ of zero planar measure.
(e) Almost everywhere equality of f and g means that there is a zero set $Z \subset \mathbb{R}$ such that if $x \notin Z$ then $f(x) = g(x)$. Apply (b), (d) to $\mathbb{R} = Z \sqcup (\mathbb{R} \setminus Z)$.
(f) Scaling the y-axis by the factor a scales planar measure correspondingly.
(g) is a matter of looking carefully at the right picture, namely Figure 125. We assume that f is integrable (otherwise (g) merely asserts that $\infty = \infty$) and define the ***f*-translation** $\mathbb{R}^2 \to \mathbb{R}^2$ as

$$T_f : (x, y) \mapsto (x, f(x) + y).$$

T_f bijects the plane to itself, and as is shown in Figure 125,

$$\mathcal{U}(f + g) = \mathcal{U}f \sqcup T_f(\mathcal{U}g). \tag{3}$$

This makes (g) an immediate consequence of the following assertion.

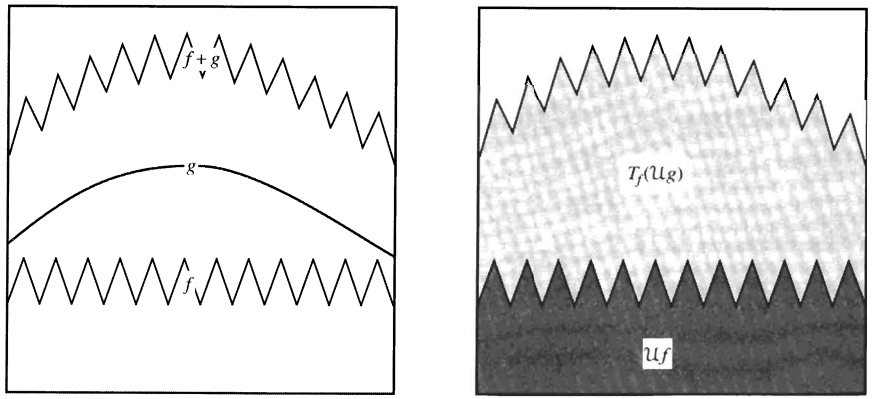


Figure 125 The undergraph of a sum.

(4) For each measurable $E \subset \mathbb{R}^2$, $T_f E$ is measurable and $m(T_f E) = mE$.

Case 1. E is a rectangle $R = (a, b) \times [0, h)$. Then $R = \mathcal{U}g$ where $g = h \cdot \chi_{(a,b)}$, and (3) gives

$$\mathcal{U}f \sqcup T_f R = \mathcal{U}(f + g) = \mathcal{U}(g + f) = R \sqcup T_g(\mathcal{U}f). \tag{5}$$

Since $(a, b) \times \mathbb{R}$ is measurable $\mathcal{U}f$ splits cleanly as $\mathcal{U}f = U_1 \sqcup U_2$ where

$$U_1 = \{(x, y) \in \mathcal{U}f : x \in (a, b)\} \qquad U_2 = \{(x, y) \in \mathcal{U}f : x \notin (a, b)\}.$$

Under T_g, U_1 is translated vertically by the amount h while U_2 stays fixed. Neither measure changes and the sets stay disjoint. Thus $m(T_g(\mathcal{U}f)) = m(\mathcal{U}f)$, and (5) becomes

$$m(\mathcal{U}f) + m(T_f R) = mR + m(\mathcal{U}f). \tag{6}$$

Since $\int f < \infty$, it is legal to subtract $m(\mathcal{U}f)$ from (6), which gives (4) when $E = R$.

Case 2. E is the rectangle $R = (a, b) \times [c, d)$. Then $T_f R = T_g R'$ where $g = f + c \cdot \chi_{(a,b)}$ and $R' = (a, b) \times [0, d - c)$. Case 1 applied to g gives (4) when $E = R$.

Case 3. E is bounded. Choose a large rectangle R that contains E. Given $\epsilon > 0$ there is a covering $\{R_k\}$ of E by rectangles such that

$$\sum_k |R_k| \leq mE + \epsilon.$$

Then $T_f E$ is covered by f-translated rectangles $T_f R_k$ and

$$m^*(T_f E) \leq \sum_k m(T_f R_k) = \sum_k |R_k| \leq mE + \epsilon.$$

Since $\epsilon > 0$ is arbitrary, $m^*(T_f E) \leq mE$. The same applies to $R \setminus E$, and so

$$\begin{aligned} &m^*(T_f E) + m^*(T_f(R \setminus E)) \\ &\leq mE + m(R \setminus E) = |R| = m(T_f R) \\ &\leq m^*(T_f E) + m^*(T_f R \setminus T_f E). \end{aligned}$$

This gives equality $m(T_f R) = m^*(T_f E) + m^*(T_f R \setminus T_f E)$, which means that $T_f E$ divides the bounded, measurable set $T_f R$ cleanly. Theorem 13 implies that $T_f E$ is measurable. Subtraction of

$$\begin{aligned} |R| &= mE + m(R \setminus E) \\ m(T_f R) &= m(T_f E) + m(T_f(R \setminus E)) \end{aligned}$$

gives

$$0 = (mE - m(T_f E)) + (m(R \setminus E) - m(T_f(R \setminus E))).$$

Both terms are ≥ 0, so they are zero, which completes the proof of (4) when E is bounded.

Case 4. E is unbounded. Break E into countably many bounded, measurable, disjoint pieces and apply Case 3 piece by piece. This completes the proof of (4), of (g), and of the theorem. □

Remark The standard proof of linearity of the Lebesgue integral is outlined in Exercise 28. It is no easier than this undergraph proof, and undergraphs at least give you a picture as guidance.

Definition The **completed undergraph** is the undergraph together with the graph itself,

$$\widehat{\mathcal{U}} f = \mathcal{U} f \cup \{(x, f(x)) : x \in \mathbb{R}\}.$$

17 Theorem *The definitions of measurability and integral are unaffected if we replace the undergraph with the completed undergraph.*

Proof Let $f : \mathbb{R} \to [0, \infty)$ be given, and set $f_n(x) = (1 + 1/n) f(x)$ for $n \in \mathbb{N}$. Then $\widehat{\mathcal{U}} f$ is the decreasing intersection $\bigcap \mathcal{U} f_n$ together with the x-axis.

Assume that $\mathcal{U} f$ is measurable. Then $\mathcal{U} f_n$ is measurable, the intersection $\bigcap \mathcal{U} f_n$ is measurable, and the x-axis is measurable (it is a zero set) so $\widehat{\mathcal{U}} f$ is measurable. Also $m(\widehat{\mathcal{U}} f) = m(\mathcal{U} f)$. For if $m(\mathcal{U} f) = \infty$, then $\widehat{\mathcal{U}} f \supset \mathcal{U} f$ implies that $m(\widehat{\mathcal{U}} f) = \infty$, while if $m(\mathcal{U} f) < \infty$, then the result follows from downward measure continuity (Theorem 5) and the fact that the x-axis is a zero set.

The converse is checked similarly. Set $g_n(x) = (1 - 1/n) f(x)$ and express $\mathcal{U} f$ as the increasing union of $\widehat{\mathcal{U}} g_n$, modulo a zero set on the x-axis. Measurability of $\widehat{\mathcal{U}} f$ implies measurability of $\widehat{\mathcal{U}} g_n$ implies measurability of $\mathcal{U} f$ and by upward measure continuity $m(\mathcal{U} f) = m(\widehat{\mathcal{U}} f)$. □

Definition Let $f_n : \mathbb{R} \to [0, \infty)$ be a sequence of functions. Its **lower envelope** and **upper envelope** sequences are

$$\underline{f}_n(x) = \inf\{f_k(x) : k \geq n\} \quad \text{and} \quad \bar{f}_n = \sup\{f_k(x) : k \geq n\}.$$

Clearly, $\underline{f}_n$ increases and $\bar{f}_n$ decreases as $n \to \infty$, and the envelopes sandwich the original sequence.

18 Theorem *If the functions f_n are measurable, then so are its envelopes.*

Proof The undergraph of the maximum of two functions is the union of their undergraphs, while the minimum is the intersection. Keeping track of strict inequality versus nonstrict inequality, we get the formulas

$$\widehat{\mathcal{U}} \underline{f}_n = \bigcap_{k \geq n} \widehat{\mathcal{U}} f_k \qquad \mathcal{U} \bar{f}_n = \bigcup_{k \geq n} \mathcal{U} f_k,$$

and measurability follows from Theorem 17. □

19 Corollary *The pointwise limit of measurable functions is measurable.*

Proof If $f_n(x) \to f(x)$ for every x (or almost every x) as $n \to \infty$ then the same is true of the envelope functions. Then $\underline{f}_n \uparrow f$ implies that $\mathcal{U} \underline{f}_n \uparrow \mathcal{U} f$ so f is measurable. □

20 Lebesgue Dominated Convergence Theorem *Suppose that $f_n : \mathbb{R} \to [0, \infty)$ is a sequence of measurable functions converging pointwise to the limit function f. If there exists a function $g : \mathbb{R} \to [0, \infty)$ whose integral is finite and which is an upper bound for all the functions f_n, $0 \leq f_n(x) \leq g(x)$, then $\int f_n \to \int f$.*

Proof Easy. See Figure 126. The upper and lower envelopes converge monotonically to f, and since $\mathcal{U} \bar{f}_n \subset \mathcal{U} g$, all have finite measure. This makes both downward and upward measure continuity valid, so we see that the measures of $\mathcal{U} \underline{f}_n$ and $\mathcal{U} \bar{f}_n$ both converge to $m(\mathcal{U} f)$. Between these measures is sandwiched $m(\mathcal{U} f_n)$, and so it also converges to $m(\mathcal{U} f)$ as $n \to \infty$. □

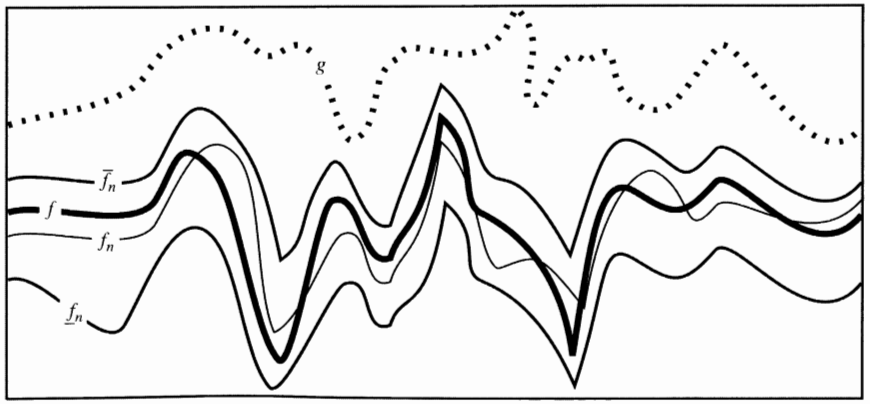


Figure 126 Lebesgue Dominated Convergence.

Remark If a dominator g with finite integral fails to exist then the assertion fails. For example the sequence of "steeple functions," shown in Figure 85 on page 204, have integral n and converge at all x to the zero function as $n \to \infty$.

21 Fatou's Lemma *Suppose that $f_n : \mathbb{R} \to [0, \infty)$ is a sequence of measurable functions and $f : \mathbb{R} \to [0, \infty)$ is their* lim inf, $f(x) = \liminf_{n\to\infty} \{f_k(x) : k \geq n\}$. *Then* $\int f \leq \liminf_{n\to\infty} \int f_n$.

Proof The assertion is really more about lim inf's than integrals. The assumption is that f is the limit of the lower envelope functions $\underline{f}_n$. Since $\underline{f}_n(x) \leq f_n(x)$, we have

$$\int \underline{f}_n \leq \int f_n.$$

The Lebesgue Monotone Convergence Theorem implies that $\int \underline{f}_n$ converges upward to $\int f$, so the integrals $\int f_n$ have a lim inf that is no smaller than $\int f$. □

Remark The inequality can be strict, as is shown by the steeple functions.

Up to now we have assumed the integrand f is nonnegative. If f takes both positive and negative values we define

$$f_+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0 \\ 0 & \text{if } f(x) < 0 \end{cases} \qquad f_-(x) = \begin{cases} -f(x) & \text{if } f(x) < 0 \\ 0 & \text{if } f(x) \geq 0. \end{cases}$$

Then $f_\pm \geq 0$ and $f = f_+ - f_-$. It is easy to see that the **total undergraph** of f,

$$\{(x, y) \in \mathbb{R}^2 : y < f(x)\}$$

is measurable if and only if $f_\pm$ are measurable. See Exercise 31. If $f_\pm$ are integrable we say that f is integrable and define its integral as

$$\int f = \int f_+ - \int f_-.$$

22 Proposition *The set of measurable functions $f : \mathbb{R} \to \mathbb{R}$ is a vector space, the set of integrable functions is a subspace, and the integral is a linear map from the latter into $\mathbb{R}$.*

The proof is left to the reader as Exercise 21.

5 Lebesgue integrals as limits

The Riemann integral is the limit of Riemann sums. There are analogous "Lebesgue sums" of which the Lebesgue integral is the limit.

Let $f : \mathbb{R} \to [0, \infty)$ be given, take a partition $Y : 0 = y_0 < y_1 < y_2 \ldots$ on the y-axis, and set

$$X_i = \{x \in \mathbb{R} : y_{i-1} \leq f(x) < y_i\}.$$

(We require that $y_i \to \infty$ as $i \to \infty$.) Assume that X_i is measurable and define the **lower and upper Lebesgue sums** as

$$\underline{L}(f, Y) = \sum_{i=1}^{\infty} y_{i-1} \cdot mX_i \qquad \bar{L}(f, Y) = \sum_{i=1}^{\infty} y_i \cdot mX_i.$$

These sums represent the measure of "Lebesgue rectangles" $X_i \times [0, y_{i-1})$ and $X_i \times [0, y_i)$ that sandwich the undergraph. Under the measurability conditions explained below, they converge to the Lebesgue integral as the mesh of Y tends to zero,

$$\underline{L}(f, Y) \uparrow \int f \qquad \text{and} \qquad \bar{L}(f, Y) \downarrow \int f.$$

Upshot Lebesgue sums are like Riemann sums, and Lebesgue integration is like Riemann integration, except that Lebesgue partitions the value axis and takes limits, while Riemann does the same on the domain axis.

The assumption that the sets X_i are measurable needs to be put in context.

Definition The function $f : \mathbb{R} \to \mathbb{R}$ is **pre-image measurable** if for each $a \in \mathbb{R}$, $f^{\text{pre}}[a, \infty) = \{x \in \mathbb{R} : a \leq f(x)\}$ is Lebesgue measurable. (It follows that the sets $X_i = f^{\text{pre}}[y_{i-1}, y_i)$ are measurable.)

This is the standard definition for measurability of a function. We will show that it is equivalent to the geometric definition that the undergraph is measurable, and then discuss when the Lebesgue sums converge to the Lebesgue integral. First we show that there is nothing special about using closed rays $[a, \infty)$.

23 Proposition *The following are equivalent conditions for pre-image measurability of $f : \mathbb{R} \to \mathbb{R}$.*

(a) The pre-image of every closed ray $[a, \infty)$ is measurable.

(b) The pre-image of every open ray (a, ∞) is measurable.

(c) The pre-image of every closed ray $(-\infty, a]$ is measurable.
(d) The pre-image of every open ray $(-\infty, a)$ is measurable.
(e) The pre-image of every half-open interval $[a, b)$ is measurable.
(f) The pre-image of every open interval (a, b) is measurable.
(g) The pre-image of every half-open interval $(a, b]$ is measurable.
(h) The pre-image of every closed interval $[a, b]$ is measurable.

Proof We show that (a) $\Rightarrow$ (b) $\Rightarrow \cdots \Rightarrow$ (h) $\Rightarrow$ (a). Since the pre-image of a union, an intersection, or a complement is the union, intersection, or complement of the pre-image, these implications are checked by taking pre-images of the following.

(a) $\Rightarrow$ (b) $(a, \infty) = \bigcup [a + 1/n, \infty)$.
(b) $\Rightarrow$ (c) $(-\infty, a] = (a, \infty)^c$.
(c) $\Rightarrow$ (d) $(-\infty, a) = \bigcup (-\infty, a - 1/n]$.
(d) $\Rightarrow$ (e) $[a, b) = (-\infty, a)^c \cap (-\infty, b)$.
(e) $\Rightarrow$ (f) $(a, b) = \bigcup [a + 1/n, b)$.
(f) $\Rightarrow$ (g) $(a, b] = \bigcap (a, b + 1/n)$.
(g) $\Rightarrow$ (h) $[a, b] = \bigcap (a - 1/n, b]$.
(h) $\Rightarrow$ (a) $[a, \infty) = \bigcup [a, a + n)$.

Remark You might expect that a measurable function is defined by the property that the pre-image of a measurable set is measurable. As is shown in Exercise 24, this is not true. The pre-image of a measurable set under a measurable function can be nonmeasurable.

24 Theorem *Pre-image measurability of $f : \mathbb{R} \to [0, \infty)$ is equivalent to measurability of $\mathcal{U}f$.*

Proof that pre-image measurability implies undergraph measurability The undergraph is

$$\mathcal{U}f = \bigcup_{r \in \mathbb{Q}} f^{\text{pre}}[r, \infty) \times [0, r),$$

and Theorem 14 asserts that the product of measurable sets is measurable. □

The converse requires a lemma.

25 Lemma *If $f : \mathbb{R} \to [0, \infty)$ is pre-image measurable and $\int f = 0$ then $f(x) = 0$ almost everywhere.*

Proof We have shown that pre-image measurability implies undergraph measurability, so $\int f$ is well defined. The product set $f^{\text{pre}}[r, \infty) \times [0, r)$ is measurable and its measure is $m(f^{\text{pre}}[r, \infty)) \cdot r = 0$. Thus $m(f^{\text{pre}}[r, \infty)) = 0$ and $f(x) = 0$ almost everywhere. □

Proof that undergraph measurability implies pre-image measurability We first assume that $|f(x)| \le M$ for all x, and $f(x) = 0$ for all $x \notin [a, b]$. Regularity implies that there is a sequence of compact sets $K_n \uparrow F \subset \mathcal{U}f$ as $n \to \infty$ such that $mF = m(\mathcal{U}f)$. Define

$$g_n(x) = \begin{cases} \max\{y : (x, y) \in K_n\} & \text{if } K_n \cap (x \times \mathbb{R}) \neq \emptyset \\ 0 & \text{if } K_n \cap (x \times \mathbb{R}) = \emptyset. \end{cases}$$

Clearly $g_n \uparrow g$ where $g \le f$ and $m(\mathcal{U}g) = m(\mathcal{U}f)$. The function g_n is upper semicontinuous. (See Exercise 26.) That is, if $x_k \to x$ as $k \to \infty$ then

$$\limsup_{k \to \infty} g_n(x_k) \le g_n(x).$$

Upper semicontinuity is equivalent to the condition that the pre-image of each open ray $(-\infty, a)$ is an open set. (See Exercise 25.) Thus g_n is pre-image measurable. The upward limit of g_n is also pre-image measurable since

$$g^{\text{pre}}(-\infty, a) = \bigcup_n g_n^{\text{pre}}(-\infty, a).$$

Because f is bounded and its support is contained in $[a, b]$, we can make the same construction from above and find a pre-image measurable function h with $f \le h$ and $m(\mathcal{U}h) = m(\mathcal{U}f)$. Thus $g \le f \le h$ and $\int g = \int f = \int h$. Linearity of the integral applies to $h = g + (h - g)$, so $\int h - g = 0$. Lemma 25 implies that $g(x) = h(x)$ almost everywhere, so f, which is sandwiched between g and f, equals a pre-image measurable function almost everywhere and hence is pre-image measurable itself.

Removing the extra boundedness hypotheses is left to the reader as Exercise 27. □

26 Corollary *If $f : \mathbb{R} \to [0, \infty)$ is measurable then $\int f = 0$ if and only if $f = 0$ almost everywhere.*

Proof Theorem 24 states that measurability implies pre-image measurability, so the assertion follows from Lemma 25. □

Now we return to Lebesgue sums. Assume that $f : \mathbb{R} \to [0, \infty]$ is measurable and define

$$f_n(x) = \begin{cases} f(x) - 1/n & \text{if } f(x) \geq 1/n \\ 0 & \text{if } f(x) < 1/n. \end{cases}$$

Clearly $f_n \uparrow f$ as $n \to \infty$, and by the Monotone Convergence Theorem $\int f_n \to \int f$. If $Y : 0 = y_0 < y_1 < \dots$ is a partition of the y-axis then the Lebesgue lower sum $\underline{L}(f, Y)$ is the integral of the function

$$f_Y(x) = \sum_{i=1}^{\infty} y_{i-1} \chi_i(x)$$

where χ_i is the characteristic function of $f^{\text{pre}}[y_{i-1}, y_i)$. If mesh $Y < 1/n$ then $f_n \leq f_Y \leq f$, and therefore $\underline{L}(f, Y)$ tends to $\int f$ as mesh $Y \to 0$.

Remark A measurable function is **simple** if it takes on only finitely many values. See Exercise 28. The lower Lebesgue sum function $\phi = \sum_{i=1}^{n} y_{i-1} \chi_{X_i}$ is a simple function, and what we have shown implies that for a nonnegative measurable function f,

$$\int f = \sup\{\int \phi : \phi \text{ is a simple function with } 0 \leq \phi \leq f\}.$$

In fact this is often how the Lebesgue integral is developed. A "pre-integral" is constructed for simple functions, and the integral of a general nonnegative measurable function is defined to be the supremum of the pre-integrals of lesser simple functions.

Lebesgue upper sums behave equally well when f is identically zero outside some set $X \subset \mathbb{R}$ with $mX < \infty$. For

$$\bar{L} - \underline{L} < \text{mesh } Y \cdot mX.$$

On the other hand, functions such as $f(x) = e^{-x^2}$ have finite integral but have every Lebesgue upper sum equal to ∞.

One way to get a satisfactory, general upper-sum/lower-sum sandwiching is to extend the concept of a partition, permitting a bi-infinite set of points $Y = \{y_i : 0 < \dots < y_{i-1} < y_i < \dots : i \in \mathbb{Z}\}$ with $y_i \to 0$ as $i \to -\infty$ and $y_i \to \infty$ as $i \to \infty$. Then

$$\liminf_{\text{mesh } Y \to 0} \bar{L}(f, Y) \to \int f.$$

See Exercise 29.

6 Italian Measure Theory

In Chapter 5 the slice method is developed in terms of Riemann integrals. Here we generalize to the Lebesgue case. The **slice** of a set $E \subset \mathbb{R}^2$ through a point x in the x-axis is

$$E_x = \{y \in \mathbb{R} : (x, y) \in E\}.$$

Similarly, the **slice** of $f : E \to \mathbb{R}$ through x is the function $f_x : E_x \to \mathbb{R}$ defined by $f_x(y) = f(x, y)$.

Remark In this section we frequently write dx and dy to indicate which variable is the integration variable.

27 Cavalieri's Principle *If $E \subset \mathbb{R}^2$ is measurable, then almost every slice E_x of E is measurable, the function $x \mapsto m_1 E_x$ is measurable, and its integral is*

$$m_2 E = \int m_1 E_x \, dx.$$

Our proof of Cavalieri's Principle involves the bisection picture used in the Bolzano-Weierstrass Theorem. A number of the form $k/2^n$ with $k, n \in \mathbb{Z}$ is **dyadic**. The vertical line $x = k/2^n$ and the horizontal line $y = \ell/2^n$ are dyadic, and so is the square with vertices $(k/2^n, \ell/2^n)$, $((k+1)/2^n, (\ell+1)/2^n)$.

28 Lemma *An open set in the plane is the union of countably many disjoint open dyadic squares together with a zero set.*

Proof Let $U \subset \mathbb{R}^2$ be open. Accept all the open unit dyadic squares that lie in U, and reject the rest. Bisect every rejected square into four equal subsquares. Accept the interiors of all these subsquares that lie in U, and reject the rest. Proceed inductively, bisecting the rejected squares, accepting the interiors of the resulting subsquares that lie in U, and rejecting the rest. In this way U is shown to be the countable union of disjoint, accepted, open dyadic squares, together with the points rejected at every step in the construction. See Figure 127.

Rejected points of U lie on horizontal or vertical dyadic lines. There are countably many such lines, each is a zero set, and so the rejected points form a zero set. □

Proof of Cavalieri's Principle Case 1. E is rectangle $R = (a, b) \times (c, d)$. Then $m_1 R_x = (d - c) \cdot \chi_{(a,b)}(x)$ and the equation

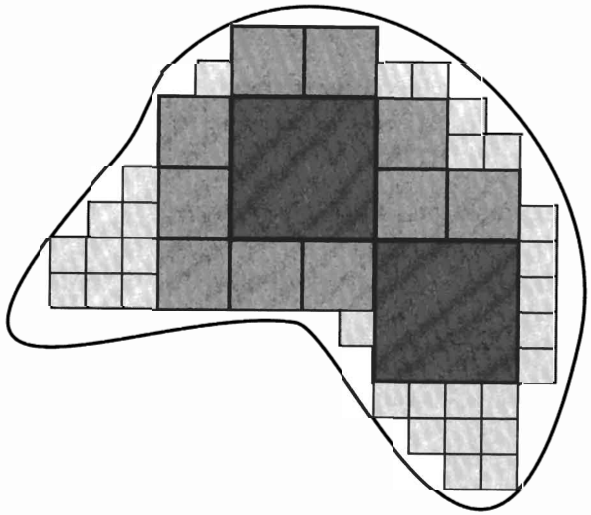


Figure 127 An open set is a countable union of dyadic squares.

$$m_2 E = (b-a)\cdot(d-c) = \int m_1 E_x \, dx \tag{7}$$

is obvious.

Case 2. E is a zero set Z. There is a sequence of coverings $\{R_{in}\}$ of Z by open rectangles such that $U_n = \bigcup_{i=1}^{\infty} R_{in}$ satisfies

$$U_1 \supset U_2 \supset \cdots \supset Z \qquad \text{and} \qquad m_2 U_n \le \sum_{i=1}^{\infty} |R_{in}| < \frac{1}{n}.$$

Define $f(x) = m_1^*(Z_x)$ and $f_n(x) = \sum_{i=1}^{\infty} m_1((R_{in})_x)$. Since $Z_x \subset \bigcup_i (R_{in})_x$, we have $0 \le f(x) \le f_n(x)$ and

$$\mathcal{U} f \subset \bigcap_{n=1}^{\infty} \mathcal{U} f_n.$$

We now show that f is measurable. Since $x \mapsto m_1((R_{in})_x)$ is a well defined measurable function of x, so is f_n, and the undergraph of f_n has measure

$$\begin{aligned} m_2(\mathcal{U} f_n) &= \int f_n(x)\, dx = \int \sum_{i=1}^{\infty} m_1((R_{in})_x)\, dx \\ &= \sum_{i=1}^{\infty} \int m_1((R_{in})_x)\, dx = \sum_{i=1}^{\infty} |R_{in}| < \frac{1}{n}. \end{aligned}$$

(We used Case 1 and the Lebesgue Monotone Convergence Theorem.) Thus $\mathcal{U} f$ is a zero set. A zero set is measurable. Therefore the function f is measurable and its integral is zero. By Corollary 26, $f(x) = 0$ almost everywhere, which implies that almost every slice of Z is a zero set and verifies Cavalieri's Principle for Z.

Case 3. E is an open set U. By Lemma 28, $U = Z \sqcup \bigsqcup S_i$ where Z is a zero set and the S_i's are open rectangles. Set $S = \bigsqcup S_i$. Cases 1, 2 and countable additivity give

$$\begin{aligned} m_2 U &= \sum_{i=1}^{\infty} m_2 S_i = \sum_{i=1}^{\infty} \int m_1((S_i)_x)\, dx = \int \sum_{i=1}^{\infty} m_1((S_i)_x)\, dx \\ &= \int m_1 S_x \, dx = \int (m_1 S_x + m_1 Z_x)\, dx = \int m_1 U_x \, dx, \end{aligned}$$

which verifies Cavalieri's Principle for an open set.

Case 4. E is a bounded measurable set. By regularity there is a sequence of bounded open sets U_n that nest down to a G_δ-set G with $G = E \sqcup Z$ and Z a zero set. Then $E_x \sqcup Z_x = G_x$, so $m_1 E_x = m_1 G_x$ almost everywhere. Also $G_x = \bigcap_{n=1}^{\infty} (U_n)_x$ and by Case 3

$$m_2 U_n = \int m_1((U_n)_x)\, dx.$$

Downward measure continuity and the Lebesgue Dominated Convergence Theorem give

$$\begin{aligned} m_2 E = m_2 G &= \lim_{n\to\infty} m_2 U_n = \lim_{n\to\infty} \int m_1((U_n)_x)\, dx \\ &= \int \lim_{n\to\infty} m_1((U_n)_x)\, dx \\ &= \int m_1 G_x \, dx = \int (m_1 E_x + m_1 Z_x)\, dx = \int m_1 E_x \, dx, \end{aligned}$$

which verifies Cavalieri's Principle for bounded measurable sets.

Case 5. E is an unbounded measurable set. Split E into bounded measurable pieces (say by intersecting it with unit dyadic squares), apply Case 4 to each piece, and add the answers. □

Cavalieri's Principle holds also in dimension 3 and higher. If $E \subset \mathbb{R}^3$, its slice through a point x on the x-axis is

$$E_x = \{(y, z) \in \mathbb{R}^2 : (x, y, z) \in E\}.$$

29 Cavalieri's Principle in 3-space *If $E \subset \mathbb{R}^3$ is measurable then almost every slice E_x of E is measurable, the function $x \mapsto m_2 E_x$ is measurable, and its integral is*

$$m_3 E = \int m_2 E_x \, dx.$$

The proof of the three-dimensional (and higher) cases is identical to the two-dimensional case. See also Appendix B of Chapter 5.

As a consequence of Cavalieri's Principle in 3-space we get the integral theorems of Fubini and Tonelli. It is standard practice to refer to the integral of a function f on $\mathbb{R}^2$ as a **double integral** and to write it as

$$\int f = \iint f(x, y)\, dxdy.$$

It is also standard to write the **iterated integral** as

$$\int \Big[\int f_x(y)\, dy \Big] dx = \int \Big[\int f(x, y)\, dy \Big] dx.$$

30 Fubini-Tonelli Theorem *If $f : \mathbb{R}^2 \to [0, \infty)$ is measurable, then almost every slice $f_x(y)$ is a measurable function of y, the function $x \mapsto \int f_x(y)\, dy$ is measurable, and the double integral equals the iterated integral,*

$$\iint f(x, y)\, dxdy = \int \Big[\int f(x, y)\, dy \Big] dx.$$

Proof The result follows from the simple observation that the slice of the undergraph is the undergraph of the slice,

$$(\mathcal{U} f)_x = \mathcal{U} f_x. \tag{8}$$

See Figure 128. For (8) implies that $m_2((\mathcal{U} f)_x) = m_2(\mathcal{U} f_x) = \int f(x, y)\, dy$, and then Cavalieri gives

$$\begin{aligned} \iint f(x, y)\, dxdy &= m_3(\mathcal{U} f) = \int \Big[m_2((\mathcal{U} f)_x) \Big] dx \\ &= \int \Big[\int f(x, y)\, dy \Big] dx. \end{aligned}$$

□

31 Corollary *When $f : \mathbb{R}^2 \to [0, \infty)$ is measurable the order of integration in the iterated integrals is irrelevant,*

$$\int \Big[\int f(x, y)\, dy \Big] dx = \iint f(x, y)\, dxdy = \int \Big[\int f(x, y)\, dx \Big] dy.$$

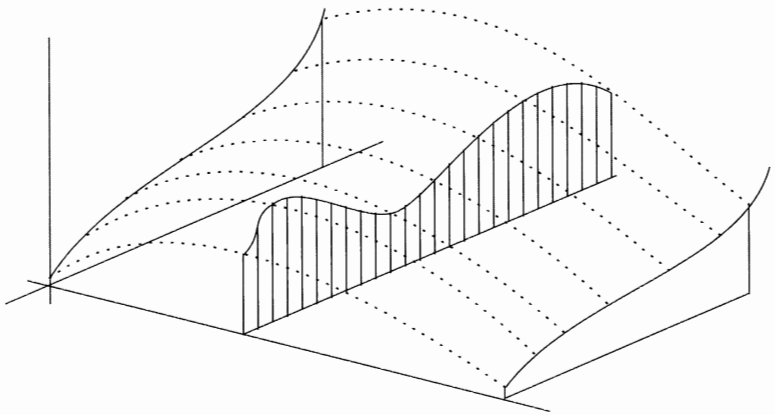


Figure 128 Slicing the undergraph.

(In particular, if one of the three integrals is finite so are the other two and all three are equal.)

Proof The difference between "x" and "y" is only notational. In contrast to the integration of differential forms, the orientation of the plane or 3-space plays no role in Lebesgue integration, so the Fubini-Tonelli Theorem applies equally to x-slicing and y-slicing, which implies that both iterated integrals equal the double integral. □

When f takes on both signs a little care must be taken to avoid subtracting ∞ from ∞.

32 Theorem *If $f : \mathbb{R}^2 \to \mathbb{R}$ is integrable (the double integral of f exists and is finite) then the iterated integrals exist and equal the double integral.*

Proof Split f into its positive and negative parts, $f = f_+ - f_-$, and apply the Fubini-Tonelli Theorem to each separately. Since the integrals are finite, subtraction is legal and the theorem follows for f. □

See Exercise 35 for an example in which trouble arises if you forget to assume that the double integral is finite.

7 Vitali coverings and density points

The fact that any open covering of a closed and bounded subset of Euclidean space reduces to a finite subcovering is certainly an important component of basic analysis. In this section, we present another covering theorem, this

time the accent being on disjointness of the sets in the subcovering rather than on finiteness. The result is used to differentiate Lebesgue integrals.

Definition A covering $\mathcal{V}$ of a set A in a metric space M is a **Vitali** covering if for each point $p \in A$ and each $r > 0$ there is $V \in \mathcal{V}$ such that $p \in V \subset M_r p$ and V is not merely the singleton set $\{p\}$. We also say that $\mathcal{V}$ is **fine** because it consists of sets that have arbitrarily small diameter.

For example if $A = [a, b]$, $M = \mathbb{R}$, and $\mathcal{V}$ consists of all intervals $[\alpha, \beta]$ with $\alpha \leq \beta$ and $\alpha, \beta \in \mathbb{Q}$ then $\mathcal{V}$ is a Vitali covering of A.

33 Vitali Covering Lemma *A Vitali covering of $A \subset \mathbb{R}^n$ by closed balls reduces to an efficient disjoint subcovering of almost all of A.*

More precisely, assume that $A \subset \mathbb{R}^n$ is bounded, $\mathcal{V}$ is a Vitali covering of A, each $V \in \mathcal{V}$ is a closed ball, and $\epsilon > 0$ is given. Then there exists a countable subcovering $\{V_k\} \subset \mathcal{V}$ of A such that

(a) The V_k are disjoint.
(b) $mU \leq m^*A + \epsilon$ where $U = \bigsqcup_{k=1}^{\infty} V_k$.
(c) $A \setminus U$ is a zero set.

Condition (b) is what we mean by $\{V_k\}$ being an **efficient** covering — the extra points covered form an ϵ-set. The sets $U_N = V_1 \sqcup \cdots \sqcup V_N$ **nearly cover** A in the sense that given $\epsilon > 0$, if N is large then U_N contains A except for an ϵ-set. After all, $U = \bigcup U_N$ contains A except for a zero set. See also Appendix C.

Boundedness of A is an unnecessary hypothesis. Also, the assumption that the sets $V \in \mathcal{V}$ are closed balls can be weakened somewhat. We discuss these improvements after the proof of the result as stated.

Proof of the Vitali Covering Lemma Given $\epsilon > 0$, there is a bounded open set $W \supset A$ such that $mW \leq m^*A + \epsilon$. Define

$$\mathcal{V}_1 = \{V \in \mathcal{V} : V \subset W\} \quad \text{and} \quad d_1 = \sup\{\operatorname{diam} V : V \in \mathcal{V}_1\}.$$

$\mathcal{V}_1$ is still a Vitali covering of A. Since W bounded, d_1 is finite. Choose $V_1 \in \mathcal{V}_1$ with $\operatorname{diam} V_1 \geq d_1/2$ and define

$$\mathcal{V}_2 = \{V \in \mathcal{V}_1 : V \cap V_1 = \emptyset\} \quad \text{and} \quad d_2 = \sup\{\operatorname{diam} V : V \in \mathcal{V}_2\}.$$

Choose $V_2 \in \mathcal{V}_2$ with $\operatorname{diam} V_2 \geq d_2/2$. In general,

$$\mathcal{V}_k = \{V \in \mathcal{V}_{k-1} : V \cap U_{k-1} = \emptyset\}$$
$$d_k = \sup\{\operatorname{diam} V : V \in \mathcal{V}_k\}$$
$$V_k \in \mathcal{V}_k \text{ has } \operatorname{diam} V_k \geq \frac{d_k}{2}$$

where $U_{k-1} = V_1 \sqcup \cdots \sqcup V_{k-1}$. This means that V_k has roughly maximal diameter among the $V \in \mathcal{V}$ that do not meet U_{k-1}. By construction, the balls V_k are disjoint and since they lie in W, $m(\bigsqcup V_k) \leq mW \leq m^*A + \epsilon$, verifying (a), (b). It remains to check (c).

If at any stage in the construction, $\mathcal{V}_k = \emptyset$, then we have covered A with finitely many V_k's, so (c) becomes trivial. We therefore assume that $V_1, V_2, \ldots$ form an infinite sequence. Additivity implies that $m(\bigsqcup V_k) = \sum mV_k$. Since all the V_k's are contained in W, the series converges. This implies that $\operatorname{diam} V_k \to 0$ as $k \to \infty$; i.e.,

$$d_k \to 0 \text{ as } k \to \infty. \tag{9}$$

For any $N \in \mathbb{N}$ we claim that

$$\bigcup_{k=N}^{\infty} 5V_k \supset A \setminus U_{N-1}. \tag{10}$$

where $5V_k$ denotes the ball V_k dilated from its center by the factor 5. (These dilated balls need not belong to $\mathcal{V}$.)

Take any $a \in A \setminus U_{N-1}$. Since U_{N-1} is compact and $\mathcal{V}_1$ is Vitali, there is a ball $B \in \mathcal{V}_1$ such that $a \in B$ and $B \cap U_{N-1} = \emptyset$. That is, $B \in \mathcal{V}_N$. Assume that (10) fails. Then, for all $k \geq N$,

$$a \notin 5V_k.$$

Therefore $B \not\subset 5V_N$. Figure 129 shows that due to the choice of V_N with roughly maximal diameter, the fact that $5V_N$ fails to contain B implies that

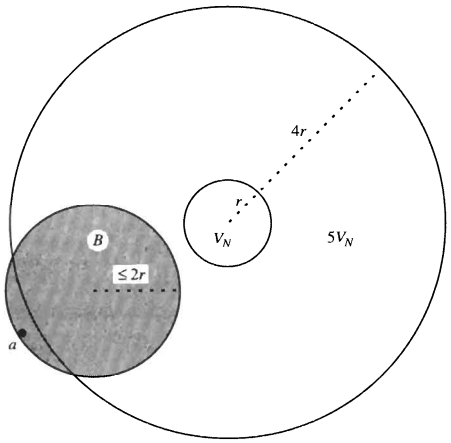


Figure 129 The unchosen ball B.

V_N is disjoint from B, so $B \in \mathcal{V}_{N+1}$. This continues for all $k > N$; namely $B \in \mathcal{V}_k$.

Aha!

B was available for choice as the next $V_k, k > N$, but it was never chosen. Therefore the chosen V_k has a diameter at least half as large as that of B. The latter diameter is fixed, but (9) states that the former diameter tends to 0 as $k \to \infty$, a contradiction. Thus (10) is true.

It is easy to see that (10) implies (c). For let $\delta > 0$ be given. Choose N so large that

$$\sum_{k=N}^{\infty} mV_k < \frac{\delta}{5^n}$$

where $n = \dim \mathbb{R}^n$. Since the series $\sum mV_k$ converges this is possible. By (10) and the scaling law $m(tE) = t^n mE$ for n-dimensional measure,

$$m^*(A \setminus U_{N-1}) \leq \sum_{k=N}^{\infty} m(5V_k) = 5^n \sum_{k=N}^{\infty} mV_k < \delta.$$

Since δ is arbitrary, $A \setminus U = \bigcap (A \setminus U_k)$ is a zero set. □

Remark A similar strategy of covering reduction appears in the proof in Chapter 2 that sequential compactness implies covering compactness. Formally, the proof is expressed in terms of the Lebesgue number of the covering, but the intuition is this: Given an open covering $\mathcal{U}$ of a sequentially compact set K, you choose a subcovering by first taking a $U_1 \in \mathcal{U}$ that covers as much of K as possible, then taking $U_2 \in \mathcal{U}$ that covers as much of the remainder of K as possible, and so on. If finitely many of these sets U_n fail to cover K then you take a sequence $x_n \in K \setminus (U_1 \cup \cdots \cup U_{n-1})$ and prove that it has no subsequence which converges in K. (The contradiction shows that in fact finitely many of the U_n you chose actually did cover K.) In short, when given a covering it is a good idea to *choose the biggest sets first.*

Removing the assumption that A is bounded presents no problem. Express an unbounded set A as $A = \bigcup A_i$ where each A_i is bounded. Given a Vitali covering $\mathcal{V}$ of A and given $\epsilon > 0$, we form an $(\epsilon/2^i)$-efficient subcovering of A_i, say $\{V_{ik} : k \in \mathbb{N}\}$. Then $\{V_{ik} : i, k \in \mathbb{N}\}$ is an ϵ-efficient subcovering of A. A further generalization involves the shapes of the sets $V \in \mathcal{V}$. See Exercise 45.

As a consequence of the Vitali Covering Lemma, we verify one of Littlewood's principles (see Appendix C):

Most points of a measurable set are like interior points.

Let $E \subset \mathbb{R}^n$ be measurable. For $x \in \mathbb{R}^n$, define the **density** of E at x as

$$\delta(x, E) = \lim_{B \downarrow x} \frac{m(E \cap B)}{mB},$$

if the limit exists, m being Lebesgue measure on $\mathbb{R}^n$ and B being a ball that shrinks down to x. Clearly, $0 \leq \delta \leq 1$. Points with $\delta = 1$ are called **density points** of E. The fraction that we're taking the limit of is the **relative measure** or **concentration** of E in B. Existence of $\delta(x, E)$ means that for each $\epsilon > 0$ there exists an $r > 0$ such that if B is any ball of radius $< r$ that contains x then the concentration of E in B differs from δ by $< \epsilon$. If we specify that the balls are centered at x, we refer to δ as the **balanced density** of E at x.

34 Lebesgue Density Theorem *Almost every $x \in E$ is a density point of E.*

Interior points of E are obviously density points of E, although sets like the irrationals or a Cantor set of positive measure have empty interior, while still having infinitely many density points.

Let E_1 denote the set of density points of E. In analogy with topological interior, it is easy to see that $(E_1)_1 = E_1$. Thus, E_1 is a kind of "measure theoretic interior" of E. Keep in mind the pathology that the boundary of an open set can have positive measure — for instance a Cantor set is the frontier of its complement and both can have positive measure. Even a Jordan curve in $\mathbb{R}^2$ can have positive Lebesgue planar measure (see Exercise 46), so this measure theoretic interior is more an analogy than a generalization.

A consequence of the Lebesgue Density Theorem is that measurable sets are not "diffuse" — a measurable subset of $\mathbb{R}$ can not meet every interval (a, b) in a set of measure $c \cdot (b-a)$ where c is a constant, $0 < c < 1$. Instead, a measurable set must be "concentrated" or "clumpy." See Exercise 42. Also, looking at the complement E^c of E, we see that almost every point $x \in E^c$ has $\delta(E, x) = 0$. Thus, almost every point of E is a density point of E and almost every point of E^c is not.

Proof of the Lebesgue Density Theorem Without loss of generality, we assume E is a bounded subset of $\mathbb{R}^n$. Take any $a, 0 \leq a < 1$, and consider

$$E_a = \{x \in E : \underline{\delta}(E, x) < a\}$$

where $\underline{\delta}$ denotes the lim inf of the concentration of E in balls B as $B \downarrow x$. We will show that E_a has outer measure zero.

By assumption, at every $x \in E_a$ there are arbitrarily small balls in which the concentration of E is $< a$. These balls form a Vitali covering of E_a and by the Vitali Covering Lemma we can select a subcovering $V_1, V_2, \ldots$ such that the V_k are disjoint, cover almost all of E_a, and nearly give the outer measure of E_a in the sense that

$$\sum_k mV_k < m^*(E_a) + \epsilon.$$

(E_a turns out to be measurable, but the Vitali Covering Lemma does not require us to know this in advance.) We get

$$m^*(E_a) = \sum_k m^*(E_a \cap V_k) < a \sum_k mV_k \leq a(m^*(E_a) + \epsilon),$$

which implies that $m^*(E_a) \leq a\epsilon/(1-a)$. Since $\epsilon > 0$ is arbitrary, $m^*(E_a) = 0$. The E_a are monotone increasing zero sets as $a \uparrow 1$. Letting $a = 1 - 1/\ell$, $\ell = 1, 2, \ldots$, we see that the union of all the E_a with $a < 1$ is also a zero set, Z. Points $x \in E \setminus Z$ have the property that as $B \downarrow x$, the lim inf of the concentration of E in B is $\geq a$ for all $a < 1$; since the concentration is always ≤ 1, this means that the limit of the concentration exists and equals 1. Hence $E \setminus Z$ is contained in E_1 and almost every point of E is a density point of E. □

8 Lebesgue's Fundamental Theorem of Calculus

In this section we write the integral of f over a set E as $\int_E f(x)\,dm$. In dimension one we write it as $\int_E f(t)\,dt$, or as $\int_\alpha^\beta f(t)\,dt$ when $E = (\alpha, \beta)$.

Definition The **average** of an integrable function $f : \mathbb{R}^n \to \mathbb{R}$ over a measurable set $E \subset \mathbb{R}^n$ with positive measure is

$$⨍_E f(x)\,dm = \frac{1}{mE} \int_E f(x)\,dm.$$

35 Theorem *If $f : \mathbb{R}^n \to \mathbb{R}$ is Lebesgue integrable, then for almost every $p \in \mathbb{R}^n$*

$$\lim_{B \downarrow p} ⨍_B f(x)\,dm = f(p),$$

where $B \downarrow p$ means that B is a ball that contains p and shrinks down to p.

Proof Case 1. $f = \chi_E$ where E is a measurable set. The average of f on B is the concentration of E in B. By the Lebesgue Density Theorem, for almost every $p \in \mathbb{R}^n$ this concentration converges to $\chi_E(p)$.

Case 2. f is integrable and nonnegative. Fix any $\alpha > 0$ and consider the set

$$A = \{p \in \mathbb{R}^n : \limsup_{B \downarrow p} \left| ⨍_B f(x)\,dm - f(p) \right| > \alpha\}.$$

It suffices to show that for each $\alpha > 0$, $A = A(\alpha)$ is a zero set. Let $\epsilon > 0$ be given. We claim that $m^*A < \epsilon$. In Section 5 we showed that for any $\epsilon > 0$ there is a Lebesgue lower sum function $\phi = \sum_{i=1} y_{i-1}\chi_{X_i}$ such that $0 \leq \phi \leq f$ and

$$\int_{\mathbb{R}^n} g(x)\,dm < \frac{\alpha\epsilon}{4} \tag{11}$$

where $g = f - \phi$.

Linearity of the integral and Case 1 imply that for almost every p, the average of ϕ on B converges to $\phi(p)$ as B shrinks to p. Thus, A differs from A' by a zero set where

$$A' = \{p \in \mathbb{R}^n : \limsup_{B \downarrow p} \left| ⨍_B g(x)\,dm - g(p) \right| > \alpha\}.$$

To get rid of the absolute values we write

$$A'_1 = \{p \in A' : g(p) > \alpha/2\}$$
$$A'_2 = \{p \in A' : \limsup_{B \downarrow p} ⨍_B g(x)\,dm > \alpha/2\},$$

and observe that $A' \subset A'_1 \cup A'_2$. Then

$$\frac{\alpha}{2} \cdot mA'_1 \leq \int_{A'_1} g(x)\,dm \leq \int_{\mathbb{R}^n} g(x)\,dm < \frac{\alpha\epsilon}{4},$$

implies that $mA'_1 \leq \epsilon/2$.

The balls B on which $⨍_B g(x)\,dm > \alpha/2$ form a Vitali covering of A'_2. The Vitali Covering Lemma gives a countable disjoint collection of these balls $\{B_i\}$ that covers A'_2. Then

$$\frac{\alpha}{2} \cdot m^*A'_2 \leq \sum_i \frac{\alpha}{2} \cdot mB_i \leq \sum_i ⨍_{B_i} g(x)\,dm \cdot mB_i \leq \int_{\mathbb{R}^n} g(x)\,dm < \frac{\alpha\epsilon}{4}.$$

Dividing through by $\alpha/2$ gives $m^* A_2' < \epsilon/2$, which completes the proof that A is a zero set, and hence completes the proof of Theorem 35 in Case 2.

Case 3. f is a general integrable function on $\mathbb{R}^n$. Write $f = f_+ - f_-$ with $f_\pm \geq 0$ and apply Case 2 to f_+ and f_-. □

36 Corollary *If $f : [a, b] \to \mathbb{R}$ is Lebesgue integrable and*

$$F(x) = \int_a^x f(t)\, dt$$

*is its **indefinite integral** then for almost every $x \in [a, b]$, $F'(x)$ exists and equals $f(x)$.*

Proof In dimension one a ball is a segment, so Theorem 35 gives

$$\frac{F(x+h) - F(x)}{h} = \fint_{[x,x+h]} f(t)\, dt \to f(x)$$

almost everywhere as $h \downarrow 0$. The same holds for $[x - h, x]$. □

Corollary 36 does not characterize indefinite integrals. Mere knowledge that a function G has a derivative almost everywhere and that its derivative is an integrable function f does *not* imply that G differs from the indefinite integral of f by a constant. The Devil's staircase function H is a counter-example. Its derivative exists almost everywhere, $H'(x)$ is almost everywhere equal to the integrable function $f(x) = 0$, and yet H does not differ from the indefinite integral of 0 by a constant. The missing ingredient is a subtler form of continuity.

Definition A function $G : [a, b] \to \mathbb{R}$ is **absolutely continuous** if for each $\epsilon > 0$ there is a $\delta > 0$ such that whenever $(\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n)$ are disjoint intervals in $[a, b]$ we have

$$\sum_{k=1}^{n} \beta_k - \alpha_k < \delta \quad \Rightarrow \quad \sum_{k=1}^{n} |G(\beta_k) - G(\alpha_k)| < \epsilon.$$

Since δ does not depend on n, it follows that if G is absolutely continuous and (α_k, β_k) is an infinite sequence of disjoint intervals in $[a, b]$ then

$$\sum_{k=1}^{\infty} \beta_k - \alpha_k < \delta \quad \Rightarrow \quad \sum_{k=1}^{\infty} |G(\beta_k) - G(\alpha_k)| \leq \epsilon.$$

37 Theorem *Let $f : [a, b] \to \mathbb{R}$ be integrable.*

(a) The indefinite integral F of f is absolutely continuous.

(b) For almost every x, $F'(x)$ exists and equals $f(x)$.

(c) If G is an absolutely continuous function and $G'(x) = f(x)$ for almost every x, then G differs from F by a constant.

As we show in the next section, the tacit assumption in (c) that $G'(x)$ exists is redundant. Theorem 37 then gives a characterization of indefinite integrals as follows.

38 Lebesgue's Main Theorem *Every indefinite integral is absolutely continuous and, conversely, every absolutely continuous function has a derivative almost everywhere and, up to a constant, is the indefinite integral of its derivative.*

Proof of Theorem 37

(a) Case 1. f is the characteristic function of a measurable set $f = \chi_E$. For any interval $(\alpha, \beta) \subset [a, b]$ we have

$$F(\beta) - F(\alpha) = \int_\alpha^\beta \chi_E(t)\, dt = m(E \cap (\alpha, \beta)).$$

Thus, given $\epsilon > 0$, we take $\delta = \epsilon$ and check that if $\{(\alpha_k, \beta_k)\}$ are disjoint intervals in $[a, b]$ with $\sum_k \beta_k - \alpha_k < \delta$ then

$$\sum_k F(\beta_k) - F(\alpha_k) = \sum_k m(E \cap (\alpha_k, \beta_k)) < \epsilon.$$

Case 2. $f \geq 0$. Given $\epsilon > 0$, choose a simple function $0 \leq \phi \leq f$ as in the proof of Lebesgue's Fundamental Theorem of Calculus (Theorem 35) such that $\int_a^b g(t)\, dt < \epsilon/2$ where $g = f - \phi$. Linearity of the integral and Case 1 applied to the indefinite integral Φ of ϕ give a δ such that if $\sum \beta_k - \alpha_k < \delta$ then $\sum_k \Phi(\beta_k) - \Phi(\alpha_k) < \epsilon/2$. Thus

$$\begin{aligned} \sum_k F(\beta_k) - F(\alpha_k) &= \sum_k \int_{\alpha_k}^{\beta_k} g(t)\, dt + \sum_k \Phi(\beta_k) - \Phi(\alpha_k) \\ &\leq \int_a^b g(t)\, dt + \frac{\epsilon}{2} < \epsilon. \end{aligned}$$

Case 3. The general integrable f. Express f as $f = f_+ - f_-$ with $f_\pm \geq 0$. If $(\alpha, \beta) \subset [a, b]$ then

$$|F(\beta) - F(\alpha)| = \left| \int_\alpha^\beta f(t)\, dt \right| \leq \int_\alpha^\beta |f(t)|\, dt = \int_\alpha^\beta f_+(t) + f_-(t)\, dt,$$

which implies that Case 3 follows from Case 2 and completes the proof of (a).

(b) This is Corollary 36: the derivative of the indefinite integral is the integrand.

(c) G is absolutely continuous and $G'(x) = f(x)$ almost everywhere. It is easy to see that the sum and difference of absolutely continuous functions are absolutely continuous. Thus $H = F - G$ is an absolutely continuous function such that $H'(x) = 0$ almost everywhere, and our goal is to show that H is constant.

Fix any $x^* \in [a, b]$. Given $\epsilon > 0$ it is enough to show that

$$\left|H(x^*) - H(a)\right| < \epsilon.$$

Absolute continuity implies that there is a $\delta > 0$ such that if $(\alpha_k, \beta_k) \subset [a, b]$ are disjoint, $k = 1, 2, \ldots$, then

$$\sum_k \beta_k - \alpha_k < \delta \quad \Rightarrow \quad \sum_k |H(\beta_k) - H(\alpha_k)| < \frac{\epsilon}{2}.$$

Fix such a δ and define

$$X = \{x \in [a, x^*] : H'(x) \text{ exists and } H'(x) = 0\}.$$

Each $x \in X$ is contained in arbitrarily small intervals $[x, x + h]$ such that

$$\left|\frac{H(x+h) - H(x)}{h}\right| < \frac{\epsilon}{2(b-a)}.$$

These intervals form a Vitali covering of X and the Vitali Covering Lemma gives a countable disjoint subcovering $\{[x_k, x_k + h_k]\}$ of almost all of X. Since $[a, x^*] \setminus X$ is a zero set, $\sum_{k=1}^{\infty} h_k = x^* - a$. Therefore there is an N such that

$$\sum_{k=1}^{N} h_k > x^* - a - \delta.$$

Since N is fixed we can re-label the indices so that $x_1 < x_2 < \cdots < x_N$. This gives a partition of $[a, x^*]$ as

$$a \le x_1 < x_1 + h_1 \le \cdots \le x_N < x_N + h_N \le x^*,$$

and we telescope $H(x^*) - H(a)$ accordingly,

$$\begin{aligned} H(x^*) - H(a) = {} & H(x^*) - H(x_N + h_N) && + H(x_N + h_N) - H(x_N) \\ & + H(x_N) - H(x_{N-1} + h_{N-1}) && + H(x_{N-1} + h_{N-1}) - H(x_{N-1}) \\ & + \cdots \\ & + H(x_2) - H(x_1 + h_1) && + H(x_1 + h_1) - H(x_1) \\ & + H(x_1) - H(a). \end{aligned}$$

The absolute values of the terms in the first column add up to at most $\epsilon/2$ since the total lengths of the intervals $(a, x_1), (x_1+h_1, x_2), \ldots, (x_N+h_N, x^*)$ is $< \delta$. The absolute value of the k^{th} term in the second column is

$$|H(x_k + h_k) - H(x_k)| < \frac{\epsilon h_k}{2(b-a)},$$

the sum of which is $< \epsilon/2$. Therefore $\left|H(x^*) - H(a)\right| < \epsilon/2 + \epsilon/2 = \epsilon$, which completes the proof that H is constant. □

9 Lebesgue's Last Theorem

The final theorem in Lebesgue's ground breaking book, *Leçons sur l'intégration*, is extremely concise and quite surprising.

39 Theorem *A monotone function has a derivative almost everywhere.*

Note that no hypothesis is made about continuity of the monotone function. Considering the fact that a monotone function $f : [a, b] \to \mathbb{R}$ has only a countable number of discontinuities, all of jump type, this may seem reasonable, but remember: the discontinuity set may be dense in $[a, b]$.

We assume henceforth that f is nondecreasing, since the non-increasing case can be handled by looking at $-f$.

Lebesgue's proof of Theorem 39 used the full power of the machinery he had developed for his new integration theory. In contrast, the proof given below is more direct and geometric. It relies on the Vitali Covering Lemma and the following form of Chebyshev's inequality from probability theory.

The **slope** of f **over** $[a, b]$ is

$$s = \frac{f(b) - f(a)}{b - a}.$$

40 Chebyshev Lemma *Assume that $f : [a, b] \to \mathbb{R}$ is nondecreasing and has slope s over $I = [a, b]$. If I contains countably many disjoint subintervals I_k and the slope of f over I_k is $\geq S > s$ then*

$$\sum_k |I_k| \leq \frac{s}{S} |I| .$$

Proof Write $I_k = [a_k, b_k]$. Since f is nondecreasing

$$f(b) - f(a) \geq \sum_k f(b_k) - f(a_k) \geq \sum_k S(b_k - a_k).$$

Thus, $s\,|I| \geq S \sum |I_k|$, and the lemma follows. □

Remark An extreme case of this situation occurs when the slope is concentrated in the three subintervals drawn in Figure 130.

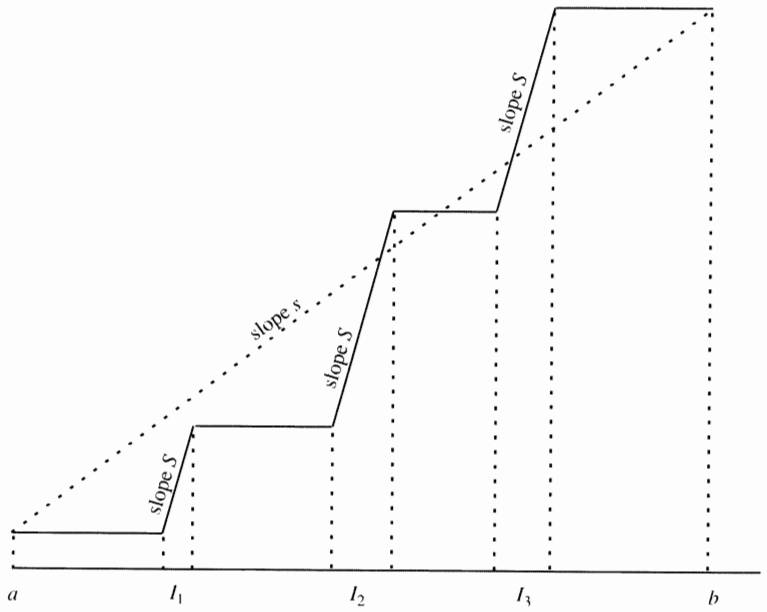


Figure 130 Chebyshev's Inequality for slopes.

Proof of Lebesgue's Last Theorem Not only will we show that $f'(x)$ exists almost everywhere, we will also show that $f'(x)$ is a measurable function of x, and

$$\int_a^b f'(x)\,dx \leq f(b) - f(a). \tag{12}$$

To estimate differentiability, one introduces upper and lower limits of slopes called derivates. If $h > 0$, $[x, x + h]$ is a **right interval** at x and $(f(x+h) - f(x))/h$ is a **right slope** at x. The lim sup of the right slopes as $h \to 0$ is called the **right maximum derivate** of f at x, $D^{\text{right max}} f(x)$, and the lim inf is the **right minimum derivate** of f at x, $D^{\text{right min}} f(x)$. Similar definitions apply to the left of x. Think of $D^{\text{right max}} f(x)$ as the steepest slope at the right of x and $D^{\text{right min}} f(x)$ as the gentlest. See Figure 131.

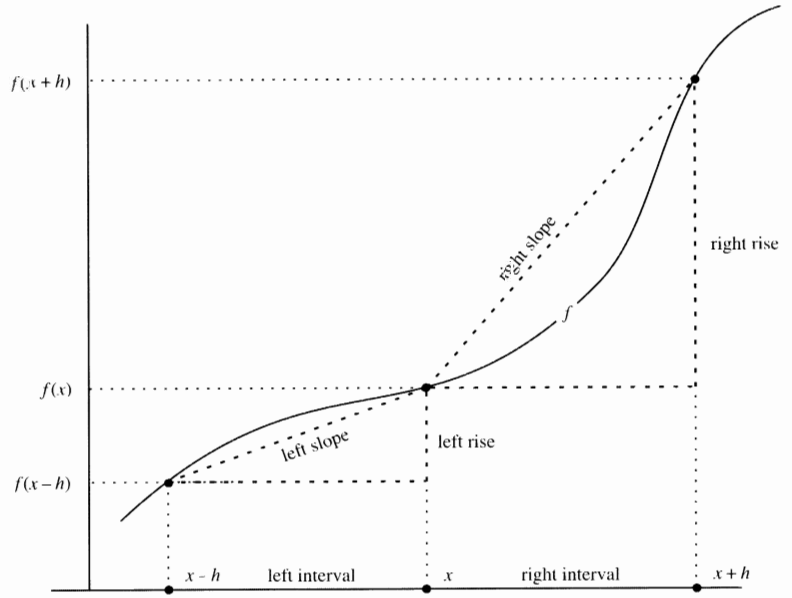


Figure 131 Left and right slopes.

The derivates exist at all points of $[a, b]$, but they can take the value ∞.

There are four derivates. We first show that two are equal almost everywhere, say the left min and the right max. Fix any $s < S$ and consider the set

$$E = E_{sS} = \{x \in [a, b] : D^{\text{left min}} f(x) < s < S < D^{\text{right max}} f(x)\}.$$

We claim that

$$m^* E = 0. \tag{13}$$

At each $x \in E$ there are arbitrarily small left intervals $[x - h, x]$ over which the slope is $< s$. These left intervals form a Vitali covering $\mathcal{L}$ of E.

(Note that the point x is not the center of its $\mathcal{L}$-interval, but rather it is an endpoint. Also, we do not know a priori that E is measurable. Luckily, Vitali permits this.) Let $\epsilon > 0$ be given. By the Vitali Covering Lemma, there are countably many disjoint left intervals $L_i \in \mathcal{L}$ that cover E, modulo a zero set, and they do so ϵ-efficiently. That is, if we write

$$L = \bigsqcup_i \operatorname{int} L_i$$

then $E \setminus L$ is a zero set and $mL \le m^*E + \epsilon$.

Every $y \in L \cap E$ has arbitrarily small right intervals $[y, y+t] \subset L$ over which the slope is $> S$. (Here it is useful that L is open.) These right intervals form a Vitali covering $\mathcal{R}$ of $L \cap E$, and by the Vitali Covering Lemma, we can find a countable number of disjoint intervals $R_j \in \mathcal{R}$ that cover $L \cap E$, modulo a zero set. Since $L \cap E = E$, modulo a zero set, $R = \bigsqcup R_j$ also covers E, modulo a zero set. By the Chebyshev Lemma,

$$m^*E \le mR = \sum_i \sum_{R_j \subset L_i} |R_j| \le \sum_i \frac{s}{S} |L_i| \le \frac{s}{S}(m^*E + \epsilon).$$

Since the inequality holds for all $\epsilon > 0$, it holds also with $\epsilon = 0$, which implies that $m^*E = 0$, and completes the proof of (13). Then

$$\{x : D^{\text{left min}} f(x) < D^{\text{right max}} f(x)\} = \bigcup_{\{(s,S) \in \mathbb{Q}\times\mathbb{Q} \,:\, s<S\}} E_{sS},$$

is a zero set. Symmetrically, $\{x : D^{\text{left min}} f(x) > D^{\text{right max}} f(x)\}$ is a zero set, and $D^{\text{left min}} f(x) = D^{\text{right max}} f(x)$ almost everywhere. Mutual equality of the other derivates, almost everywhere, is checked in the same way. See Exercise 49.

So far we have shown that for almost every $x \in [a, b]$, the derivative of f at x exists, although it may equal ∞. Infinite slope is not really acceptable, and that is the purpose of (12) — for an integrable function takes on a finite value at almost every point.

The proof of (12) uses a cute trick reminiscent of the traveling secant method. First extend f from $[a, b]$ to $\mathbb{R}$ by setting $f(x) = f(a)$ for $x < a$ and $f(x) = f(b)$ for $x > b$. Then define $g_n(x)$ to be the slope of the secant from $(x, f(x))$ to $(x + 1/n, f(x + 1/n))$. That is,

$$g_n(x) = \frac{f(x + 1/n) - f(x)}{1/n} = n(f(x + 1/n) - f(x)).$$

See Figure 132. Since f is almost everywhere continuous, it is measurable, and so is g_n. For almost every x, $g_n(x)$ converges to $f'(x)$ as $n \to \infty$. Hence f' is measurable, and clearly $f' \ge 0$. Fatou's Lemma gives

$$\int_a^b f'(x)\,dx = \int_a^b \liminf_{n\to\infty} g_n(x)\,dx \le \liminf_{n\to\infty} \int_a^b g_n(x)\,dx.$$

The integral of g_n is

$$\int_a^b g_n(x)\,dx = n\int_b^{b+1/n} f(x)\,dx \;-\; n\int_a^{a+1/n} f(x)\,dx.$$

The first integral equals $f(b)$ since we set $f(x) = f(b)$ for $x > b$. The second integral is at least $f(a)$ since f is nondecreasing. Thus

$$\int_a^b g_n(x)\,dx \le f(b) - f(a),$$

which completes the proof of (12). As remarked before, since the integral of f' is finite, $f'(x) < \infty$ for almost all x, and hence f is differentiable (with finite derivative) almost everywhere. □

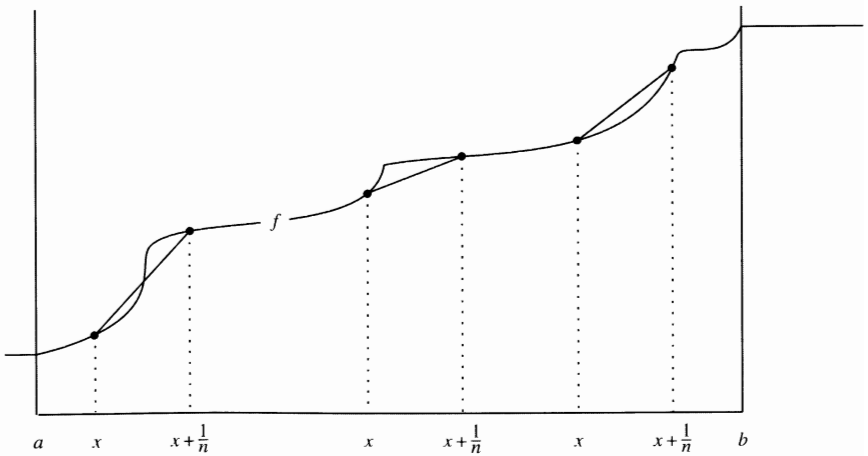


Figure 132 $g_n(x)$ is the slope of the right secant at x.

41 Corollary *A Lipschitz function is almost everywhere differentiable.*

Proof Suppose that $f : [a, b] \to \mathbb{R}$ is Lipschitz with Lipschitz constant L. Then, for all $x, y \in [a, b]$,

$$|f(y) - f(x)| \le L\,|y - x|\,.$$

The function $g(x) = f(x) + Lx$ is nondecreasing. Thus, g' exists almost everywhere and so does $f' = g' - L$. □

Remark Corollary 41 remains true for a Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, it is Rademacher's Theorem, and the proof is much harder.

Definition The **variation** of a function $f : [a, b] \to \mathbb{R}$ over a partition $X : a = x_0 < \cdots < x_n = b$ is the sum $\sum_{k=1}^{n} |\Delta_k f|$ where $\Delta_k f = f(x_k) - f(x_{k-1})$. The supremum of the variations over all partitions X is the **total variation** of f. If the total variation of f is finite, f is said to be a function of **bounded variation**.

42 Theorem *A function of bounded variation is almost everywhere differentiable.*

Proof Up to an additive constant, a function of bounded variation can be written as the difference $f(x) = P(x) - N(x)$ where

$$P(x) = \sup\{\sum_k \Delta_k f : a = x_0 < \cdots < x_n = x \text{ and } \Delta_k f \ge 0\}$$

$$N(x) = -inf\{\sum_k \Delta_k f : a = x_0 < \cdots < x_n = x \text{ and } \Delta_k f < 0\}.$$

See Exercise 52. The functions P and N are monotone nondecreasing, so for almost every x, $f'(x) = P'(x) - N'(x)$ exists and is finite. □

43 Theorem *An absolutely continuous function is of bounded variation.*

Proof Assume that $F : [a, b] \to \mathbb{R}$ is absolutely continuous and take $\epsilon = 1$. There is a $\delta > 0$ such that if (α_k, β_k) are disjoint intervals in $[a, b]$ then

$$\sum_k \beta_k - \alpha_k < \delta \quad \Rightarrow \quad \sum_k |F(\beta_k) - F(\alpha_k)| < 1.$$

Fix a partition D of $[a, b]$ with N subintervals of length $< \delta$. For any partition X of $[a, b]$ we claim that $\sum_k |\Delta_k f| < N$. We may assume that X contains D, since adding points to a partition increases the sum $\sum |\Delta_k f|$. Then

$$\sum_X |\Delta_k F| = \sum_{X_1} |\Delta_k F| + \cdots + \sum_{X_N} |\Delta_k F|$$

where X_j refers to the subintervals of X that lie in the j^{th} subinterval of D. The subintervals in X_j have total length $< \delta$, so the variation of F over them is < 1, and the total variation of F is $< N$. □

44 Corollary *An absolutely continuous function is almost everywhere differentiable.*

Proof Absolute continuity implies bounded variation, which implies almost everywhere differentiability. □

As mentioned in Section 8, Theorem 37 plus Corollary 44 express Lebesgue's Main Theorem,

Indefinite integrals are absolutely continuous and every absolutely continuous function has a derivative almost everywhere of which it is the indefinite integral.

Appendix A: Translations and Nonmeasurable sets

If $t \in \mathbb{R}$ is fixed, **t-translation** is the mapping $x \mapsto x + t$. It is a homeomorphism $\mathbb{R} \to \mathbb{R}$. Think of the circle C as $\mathbb{R}$ modulo $\mathbb{Z}$. That is, you identify any x with $x + n$ for $n \in \mathbb{Z}$. Equivalently, you take the unit interval $[0, 1]$ and you identify 1 with 0. Then t-translation becomes rotation by the angle $2\pi t$, and is denoted as $R_t : C \to C$. If t is rational this rotation is **periodic**, i.e., for some $n \ge 1$, the n^{th} **iterate** of R, $R^n = R \circ \cdots \circ R$, is the identity map $C \to C$. In fact the smallest such n is the denominator when $t = m/n$ is expressed in lowest terms. On the other hand, if t is irrational then $R = R_t$ is nonperiodic; every **orbit** $O(x) = \{R^k(x) : k \in \mathbb{Z}\}$ is denumerable and dense in C.

45 Theorem *Let $R = R_t$ with t irrational. If $P \subset C$ contains exactly one point of each orbit of R, then P is nonmeasurable.*

Proof The orbits of R are disjoint sets, there are uncountably many of them, and they divide the circle as $C = \bigsqcup_{n\in\mathbb{Z}} R^n(P)$. Translation preserves outer measure, measurability, and measure. So does rotation. Can P be

measurable? No, because if it is measurable with positive measure we would get

$$m(C) = \sum_{n=-\infty}^{\infty} m(R^n P) = \infty,$$

a contradiction, while if $mP = 0$ then $mC = \sum_{n=-\infty}^{\infty} m(R^n P) = 0$, which contradicts the fact that $m[0, 1) = 1$. □

But does P exist? The Axiom of Choice states that given any family of non-empty disjoint sets there exists a set that contains exactly one element from each set. So, if you accept the Axiom of Choice, you apply it to the family of R-orbits, and you get an example of a nonmeasurable set; while if you don't accept the Axiom of Choice you're out of luck.

To increase the pathology of P, we discuss translations in more depth below.

46 Theorem *If $E \subset \mathbb{R}$ is measurable and has positive measure, then there exists a $\delta > 0$ such that for all $t \in (-\delta, \delta)$, the t-translate of E meets E.*

47 Lemma *If $F \subset (a, b)$ is measurable and disjoint from its t-translate, then*

$$2mF \leq (b - a) + |t|.$$

Proof F and its t-translate have equal measure, so if they do not intersect, then their total measure is $2mF$, and any interval that contains them must have length $\geq 2mF$. If $t > 0$ then $(a, b + t)$ contains F and its t-translate, while if $t < 0$ then $(a + t, b)$ contains them. The length of the interval in either case is $(b - a) + |t|$. □

Proof of Theorem 46 By the Lebesgue Density Theorem (Theorem 34) E has lots of density points, so we can find an interval (a, b) in which E has concentration $> 1/2$. Call $F = E \cap (a, b)$. Then $mF > (b - a)/2$. By Lemma 47 if $|t| < 2mF - (b - a)$ then the t-translate of F meets F, so the t-translate of E meets E, which is what the theorem asserts. □

Now we return to the nonmeasurable set P discussed in Theorem 45. It contains exactly one point from each orbit of R, R being rotation by an irrational t. Set

$$A = \bigcup_{k \in \mathbb{Z}} R^{2k} P \qquad B = \bigcup_{k \in \mathbb{Z}} R^{2k+1} P.$$

The sets A, B are disjoint, their union is the circle, and R interchanges them. Since R preserves outer measure $m^* A = m^* B$.

The composite $R^2 = R \circ R$ is rotation by $2t$, also an irrational number. Let $\epsilon > 0$ be given. Since the orbit of 0 under R^2 is dense, there is a large integer k with

$$\left|R^{2k}(0) - (-t)\right| < \epsilon.$$

For R^{2k} is the k^{th} iterate of R^2. Thus $\left|R^{2k+1}(0)\right| < \epsilon$ so R^{2k+1} is a rotation by $< \epsilon$. Odd powers of R interchange A and B, so odd powers of R translate A and B off themselves. It follows from Theorem 46 that A and B contain no subsets of positive measure. Their inner measures are zero.

The general formula $mC = m_* A + m^* B$ implies that $m^* B = 1$. Thus we get an extreme type of nonmeasurability expressed in the next theorem.

48 Theorem *The circle, or equivalently $[0, 1)$, splits into two nonmeasurable, disjoint subsets that both have inner measure zero and outer measure one.*

Appendix B: The Banach-Tarski Paradox

If the example in the preceding appendix does not disturb you enough, here is a much worse one. You can read about it in Stan Wagon's book, *The Banach-Tarski Paradox*. Many other paradoxes are discussed there too.

The solid unit ball in three-space can be divided into five disjoint sets, $A_1, \ldots, A_5$, and the A_i can be moved by rigid motions to new disjoint sets A_i' whose union is two disjoint unit balls. The Axiom of Choice is fundamental in the construction, as is dimensionality greater than two. The sets A_i are nonmeasurable.

Think of this from an alchemist's point of view. A one inch gold ball can be cut into five disjoint pieces and the pieces rigidly re-assembled to make two one inch gold balls. Repeating the process would make you very rich.

Appendix C: Riemann integrals as undergraphs

The geometric description of the Lebesgue integral as the measure of the undergraph has a counterpart for Riemann integrals.

49 Theorem *A function $f : [a, b] \to [0, M]$ is Riemann integrable if and only if the Lebesgue measure of the closure and interior of its undergraph are equal. Equivalently the boundary of its undergraph is a Lebesgue zero set.*

Proof Assume that f is Riemann integrable. Given $\epsilon > 0$, a fine partition of the $[a, b]$ gives upper and lower sums, $U = \sum M_i \Delta x_i$ and $L = \sum m_i \Delta x_i$, which differ by $< \epsilon$. As usual, $\Delta x_i = x_i - x_{i-1}$, while m_i and M_i are the infimum and supremum of $f(x)$ as x varies in $[x_{i-1}, x_i]$. Geometrically this means that the graph of f is covered by rectangles $R_i = X_i \times Y_i$ where $X_i = [x_{i-1}, x_i]$ and $Y_i = [m_i, M_i]$. The open rectangles $(x_{i-1}, x_i) \times (0, m_i)$ are disjoint, lie in the interior of the undergraph, and have total area L, while the closed rectangles $[x_{i-1}, x_i] \times [0, M_i]$ cover the closure of the undergraph and have total area U. Since $U - L < \epsilon$ and $\epsilon > 0$ is arbitrary, the interior and closure of the undergraph have equal measure.

Conversely, assume that $m(\text{interior}\,\mathcal{U}f) = m(\text{closure}\,\mathcal{U}f)$ and let $\epsilon > 0$ be given. By regularity of planar Lebesgue measure, there are a compact subset K of the interior of the undergraph and an open set V that contains the closure of the undergraph such that

$$m(V \setminus K) < \epsilon.$$

For $x \in [a, b]$ set $k(x) = \max\{y : (x, y) \in K\}$, or if no such y exists set $k(x) = 0$. The segment

$$K(x) = \{(x, t) : 0 < t \leq k(x)\}$$

lies in the interior of the undergraph. Otherwise, there would exist a sequence (x_n, y_n) that converges to some $(x, t) \in K(x)$ and (x_n, y_n) does not lie in the undergraph. Neither does the higher point $(x_n, y_n + (k(x) - t)$. But the latter tends to $(x, k(x))$, which contradicts the fact that $(x, k(x)) \in K \subset \text{interior}\,\mathcal{U}f$. Since $K(x)$ is contained in the interior of the undergraph, there is a $\delta = \delta(x) > 0$ such that the strip $(x - \delta, x + \delta) \times (0, k(x) + \delta)$ is contained in it too. These strips cover K, K is compact, and so finitely many of them cover K. The strips then refine to rectangles whose total area is less than or equal to a lower sum L with $mK \leq L$. See Figure 133.

Similarly, the open set V contains strips that cover the closure of the undergraph and they refine to rectangles whose total area is greater than or equal to an upper sum U with $U \leq mV$. Hence $U - L < \epsilon$, which implies Riemann integrability. □

50 Corollary *If f is Riemann integrable, then it is Lebesgue integrable and the two integrals are equal.*

Proof Since

$$\text{interior}\,\mathcal{U}f \subset \mathcal{U}f \subset \text{closure}\,\mathcal{U}f,$$

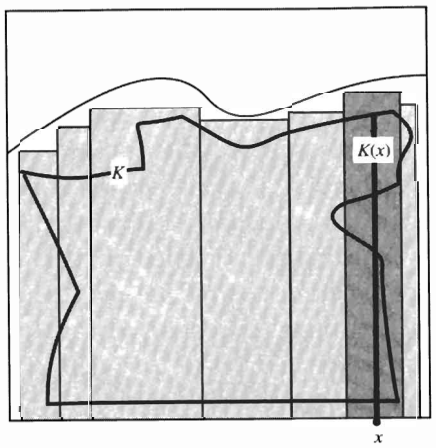


Figure 133 The strips that cover K.

equality of the measures of its interior and closure implies that $\mathcal{U}f$ is measurable, and it shares their common measure. Since the Lebesgue integral of f is equals $m(\mathcal{U}f)$ the proof is complete. □

Appendix D: Littlewood's Three Principles

In the following excerpt from his book on complex analysis, *Lectures on the Theory of Functions*, J.E. Littlewood seeks to demystify Lebesgue theory. It owes some of its popularity to its prominence in Royden's classic text, *Real Analysis*.

> The extent of knowledge [of real analysis] required is nothing like as great as is sometimes supposed. There are three principles, roughly expressible in the following terms: Every (measurable) set is nearly a finite sum of intervals; every function (of class L^λ) is nearly continuous; and every convergent sequence of functions is nearly uniformly convergent. Most of the results of the present section are fairly intuitive applications of these ideas, and the student armed with them should be equal to most occasions when real variable theory is called for. If one of the principles would be the obvious means to settle a problem if it were "quite" true, it is natural to ask if

> the "nearly" is near enough, and for a problem that is actually soluble it generally is.†

Littlewood's first principle expresses the regularity of Lebesgue measure. Given $\epsilon > 0$, a measurable $S \subset [a, b]$ contains a compact subset covered by finitely many intervals, whose union differs from S by a set of measure $< \epsilon$. In that sense, S is **nearly** a finite union of intervals. I like very much Littlewood's choice of the term "nearly," meaning "except for an ϵ-set," to contrast with "almost," meaning "except for a zero set."

Littlewood's second principle refers to "functions of class L^{λ}," although he might better have said "measurable functions." He means that if you have a (measurable) function and you are given $\epsilon > 0$, you can discard an ϵ-set from its domain of definition and the result is a continuous function. This is Lusin's Theorem: a measurable function is **nearly continuous**.

Littlewood's third principle concerns a sequence of (measurable) functions that converges almost everywhere to a limit. Except for an ϵ-set the convergence is uniform, which is Egoroff's Theorem: almost everywhere convergence implies **nearly uniform** convergence.

Proofs of Egoroff's and Lusin's theorems are outlined in Exercises 54 and 57.

Appendix E: Roundness

The density of a set E at x is the limit, if it exists, of the concentration of E in a ball B as $B \downarrow x$. What if you used a cube instead of a ball? Or an ellipsoid? Would it matter? The answer is "somewhat."

Let us say that a neighborhood U of x is **K-quasi-round** if it can be sandwiched between balls $B \subset U \subset B'$ with diam $B' \leq K$ diam B. A ball is 1-quasi-round, while a square is $\sqrt{2}$-quasi-round.

It is not hard to check that if x is a density point with respect to balls then it also a density point with respect to K-quasi-round neighborhoods of x, provided that K is fixed as the neighborhoods shrink to x. See Exercise 45. When the neighborhoods are not quasi-round, the density point analysis becomes marvelously complicated. See Falconer's book, *The Geometry of Fractal Sets*.

† Reprinted from *Lectures on the Theory of Functions* by J.E. Littlewood (1994) by permission of Oxford University Press.

Appendix F: Money

Riemann and Lebesgue walk into a room and find a table covered with hundreds of U.S. coins. (Well ...) How much money is there?

Riemann solves the problem by taking the coins one at a time and adding their values as he goes. As he picks up a penny, a nickel, a quarter, a dime, a penny, etc., he counts: "one cent, 6 cents, 31 cents, 41 cents, 42 cents, etc." The final number is Riemann's answer.

In contrast, Lebesgue first sorts the coins into piles of the same value (partitioning the value axis and taking pre-images); he then counts each pile (applying counting measure); and he sums the six terms, "value v times number of coins with value v," and that is his answer.

Lebesgue's answer and Riemann's answer are of course the same number. It is their methods of calculating that number which are different.

Now imagine that *you* walk into the room and behold this coin-laden table. Which method would you actually use to find out how much money there is — Riemann's or Lebesgue's? This amounts to the question: Which is the "better" integration theory? As an added twist, suppose you have only sixty seconds to make a good guess. What would you do then?

Suggested Reading

There are many books on more advanced analysis and topology. Among my favorites in the "not too advanced" category are these.

1. Kenneth Falconer, *The Geometry of Fractal Sets*.
   Here you should read about the Kakeya problem: how much area is needed to reverse the position of a unit needle in the plane by a continuous motion? Falconer also has a couple of later books on fractals that are good.
2. Thomas Hawkins, *Lebesgue's Theory of Integration*.
   You will learn a great deal about the history of Lebesgue integration and analysis around the turn of the last century from this book, including the fact that many standard attributions are incorrect. For instance, the Cantor set should be called the Smith set; Vitali had many of the ideas credited solely to Lebesgue, etc. Hawkins' book is a real gem.
3. John Milnor, *Topology from the Differentiable Viewpoint*.
   Milnor is one of the clearest mathematics writers and thinkers of the twentieth century. This is his most elementary book, and it is only seventy six pages long.
4. James Munkres, *Topology, a First Course*.
   This is a first year graduate text that deals with some of the same material you have been studying.
5. Robert Devaney, *An Introduction to Chaotic Dynamical Systems*.
   This is the book you should read to begin studying mathematical dynamics. It is first rate.

One thing you will observe about all these books — they use pictures to convey the mathematical ideas. Beware of books that don't.

Bibliography

1. Ralph Boas, *A Primer of Real Functions*, The Mathematical Association of America, Washington DC, 1981
2. Andrew Bruckner, *Differentiation of Real Functions*, Lecture Notes in Mathematics, Springer-Verlag, New York, 1978.
3. John Burkill, *The Lebesgue Integral*, Cambridge University Press, London, 1958.
4. Paul Cohen, *Set Theory and the Continuum Hypothesis*, Benjamin, New York, 1966.
5. Robert Devaney, *An Introduction to Chaotic Dynamical Systems*, Benjamin Cummings, Menlo Park, CA, 1986.
6. Jean Dieudonné, *Foundations of Analysis*, Academic Press, New York, 1960.
7. Kenneth Falconer, *The Geometry of Fractal Sets*, Cambridge University Press, London, 1985.
8. Russell Gordon, *The Integrals of Lebesgue, Denjoy, Perron, and Henstock*, The American Mathematical Society, Providence, RI, 1994.
9. Fernando Gouvêa, *p-adic Numbers*, Springer-Verlag, Berlin, 1997.
10. Thomas Hawkins, *Lebesgue's Theory of Integration*, Chelsea, New York, 1975.
11. George Lakoff, *Where Mathematics Comes From*, Basic Books, New York, 2000.
12. Edmund Landau, *Foundations of Analysis*, Chelsea, New York, 1951.
13. Henri Lebesgue, *Leçons sur l'intégration et la recherche des fonctions primitives*, Gauthiers-Villars, Paris, 1904.
14. John Littlewood, *Lectures on the Theory of Functions*, Oxford University Press, Oxford, 1944.
15. Ib Madsen and Jørgen Tornehave, *From Calculus to Cohomology*, Cambridge University Press, Cambridge, 1997.
16. Jerrold Marsden and Alan Weinstein, *Calculus III*, Springer-Verlag, New York, 1998.
17. Robert McLeod, *The Generalized Riemann Integral*, The Mathematical Association of America, Washington DC, 1980.
18. John Milnor, *Topology from the Differentiable Viewpoint*, Princeton University Press, Princeton, 1997.
19. Edwin Moise, *Geometric Topology in Dimensions 2 and 3*, Springer-Verlag, New York, 1977.
20. James Munkres, *Topology, a First Course*, Prentice Hall, Englewood Cliffs, NJ, 1975.

21. Murray Protter and Charles Morrey, *A First Course in Real Analysis*, Springer-Verlag, New York, 1991.
22. Dale Rolfsen, *Knots and Links*, Publish or Perish, Berkeley, 1976.
23. Halsey Royden, *Real Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
24. Walter Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, New York, 1976.
25. James Stewart, *Calculus with Early Transcendentals*, Brooks Cole, New York, 1999.
26. Arnoud van Rooij and Wilhemus Schikhof, *A Second Course on Real Functions*, Cambridge University Press, London, 1982.

Exercises

1. If $t \in \mathbb{R}$ is fixed show that the linear outer measure of A and its t-translate $t+A = \{t+a : a \in A\}$ are equal. What is the corresponding assertion in the plane?
2. If $t \geq 0$ is fixed and $A \subset \mathbb{R}$ is given, show that $m^*(tA) = t \cdot m^*A$ where $tA = \{ta : a \in A\}$ is the t-dilation of A. State the corresponding assertion in the plane.
3. Give a shorter proof of Theorem 2 in dimension one as follows:
   (a) If $I = [a, b]$ is covered by finitely many open intervals $I_k = (a_k, b_k)$, none of which contains I, show that there exists $c \in I$ such that $[a, c]$ and $[c, b]$ are covered by fewer than k of the intervals I_k.
   (b) Use (a) and induction on k to show that $|I| \leq m^*I$.
4. (a) Show that the definition of linear outer measure is unaffected if we demand that the intervals I_k in the coverings be closed instead of open.
   (b) Why does this imply that the middle-thirds Cantor set has linear outer measure zero?
   (c) Show that the definition of linear outer measure is unaffected if we drop all openness/closedness requirements on the intervals I_k in the coverings.
   (d) What about planar outer measure? Specifically, what if we demand that the rectangles be squares?
5. In analogy with intervals and rectangles, formulate a definition of an n-dimensional box R in n-space.
   (a) What is the natural definition of the volume of R?
   (b) What should be the outer measure of $A \subset \mathbb{R}^n$?
   (c) Check the outer measure axioms for your definition.
   (d) Prove that the outer measure of a box equals its volume.
6. A line in the plane that is parallel to one of the coordinate axes is a planar zero set because it is the Cartesian product of a point (which is a linear zero set) and $\mathbb{R}$.
   (a) What about a line that is not parallel to a coordinate axis?
   (b) What is the situation in higher dimensions?
7. Prove that every closed set in $\mathbb{R}$ or $\mathbb{R}^n$ is a G_δ-set. Does it follow at once that every open set is an F_σ-set? Why?
8. The **hull** of $A \subset \mathbb{R}$ or $A \subset \mathbb{R}^n$ is a smallest measurable set H that contains A, where "smallest" means that for any measurable $E \supset A$,

$H \setminus E$ is a zero set. The **kernel** of A is a largest (in the corresponding sense) measurable set K contained in A.

(a) Show that a hull of A exists, that it can be taken to be a G_δ-set, and that it is unique up to a zero set.

(b) Show that a kernel of A exists, that it can be taken to be an F_σ-set, and that it is unique up to a zero set.

(c) Show that A is measurable if and only if its hull and kernel differ by a zero set.

(d) If H is the hull of A, show that H^c is the kernel of A^c.

9. Complete the proofs of Theorems 11and 14 in the unbounded case. [Hint: How can you break an unbounded set into countably many disjoint bounded pieces?]
10. Show that inner measure is translation invariant. How does it behave under dilation?
11. Countable intersections of F_σ-sets are called $F_{\sigma\delta}$-sets. Find an $F_{\sigma\delta}$-set that is neither a G_δ-set nor an F_σ-set.
12. How does Theorem 13 fail for unbounded sets?
13. Generalize Theorem 14 to $\mathbb{R}^n$ as follows. If $A \subset \mathbb{R}^k$ and $B \subset \mathbb{R}^\ell$ are measurable show that $A \times B$ is a measurable subset of $\mathbb{R}^{k+\ell}$ and

$$m_{k+\ell}(A \times B) = m_k A \cdot m_\ell B.$$

[Hint: Generalize Lemma 28 to n dimensions and imitate the proof of Theorem 14.]

*14. Generalize the geometric undergraph characterization of Riemann integrability that appears in Appendix B to the case in which the integrand is a function of several variables.
15. Observe that under Cartesian products, measurable and nonmeasurable sets act like odd and even integers respectively.

(a) Which theorem asserts that the product of measurable sets is measurable? (Odd times odd is odd.)

(b) Is the product of nonmeasurable sets nonmeasurable? (Even times even is even.)

(c) Is the product of a nonmeasurable set and a measurable set having non-zero measure always nonmeasurable? (Even times odd is even.)

(d) Zero sets are special. They correspond to the number zero, an odd number in this imperfect analogy. (Zero times anything is zero.)

16. The **outer (Jordan) content** of a set $A \subset \mathbb{R}$ is the infimum of the total lengths of *finite* coverings of A by open intervals,

$$J^*A = \inf\{\sum_{k=1}^{n} |I_k| : \text{ each } I_k \text{ is an open interval and } A \subset \bigcup_{k=1}^{n} I_k\}.$$

The corresponding definition of outer content in the plane or n-space substitutes boxes for intervals.

(a) Show that outer content satisfies

(i) $J^*(\emptyset) = 0$.

(ii) If $A \subset B$ then $J^*A \leq J^*B$.

(iii) If $A = \bigcup_{k=1}^{n} A_k$ then $J^*A \leq \sum_{k=1}^{n} J^*A_k$.

(b) (iii) is called **finite subadditivity**. Find an example of a set $A \subset [0, 1]$ such that $A = \bigcup_{k=1}^{\infty} A_k$, $J^*A_k = 0$ for all k, and $J^*A = 1$, which shows that finite subadditivity does not imply countable subadditivity, and that J^* is not an outer measure.

(c) Why is it clear that $m^*A \leq J^*A$, and that if A is compact then $mA = J^*A$? What about the converse?

(d) Show that the requirement that the intervals in the covering of A be open is irrelevant.

17. Prove that

$$J^*A = J^*\bar{A} = mA$$

where $\bar{A}$ is the closure of A.

18. If A, B are compact show that

$$J^*(A \cup B) + J^*(A \cap B) = J^*A + J^*B.$$

[Hint: Is the formula true for Lebesgue measure? Use Exercise 17.]

19. The **inner (Jordan) content** of a subset A of an interval I is

$$J_*A = |I| - J^*(I \setminus A).$$

(a) Show that

$$J_*A = m(\text{interior } A).$$

(b) A set with equal inner and outer content is said to **have content**. Infer from Theorem 49 that a bounded function is Riemann integrable if and only if its undergraph has content.

20. Prove that measurability of $E \subset \mathbb{R}$ is equivalent to (a), to (b), and to (c).

(a) E cleanly divides every measurable subset of $\mathbb{R}$.
(b) E cleanly divides every open subset of $\mathbb{R}$.
(c) E cleanly divides every interval $[a, b] \subset \mathbb{R}$.
(d) Generalize (a), (b), (c) to the plane and $\mathbb{R}^n$.

21. Prove Proposition 22.
22. Extend Proposition 23 by showing that pre-image measurability is equivalent to
    (a) The pre-image of every G_δ-set is measurable
    (b) The pre-image of every F_σ-set is measurable.
23. Show that every $S \subset \mathbb{R}$ with $m^*S > 0$ contains a nonmeasurable subset. [Hint: Theorem 48.]
24. Consider the Devil's ski slope homeomorphism $h : [0, 1] \to [0, 2]$ that sends the standard Cantor set C to a fat Cantor set F of measure 1.
    (a) Does C contain a nonmeasurable set?
    (b) Does F contain a nonmeasurable set?
    (c) Is h^{-1} measurable?
    (d) If $E \subset [0, 1]$ is measurable, is its pre-image under h^{-1} always measurable?
25. A function $f : M \to \mathbb{R}$ is **upper semicontinuous** if

$$\lim_{k\to\infty} x_k = x \quad \Rightarrow \quad \limsup_{k\to\infty} f(x_k) \le f(x).$$

(M can be any metric space.)
    (a) Draw a graph of an upper semicontinuous function that is not continuous.
    (b) Show that upper semicontinuity is equivalent to the requirement that for every open ray $(-\infty, a)$, $f^{\text{pre}}(-\infty, a)$ is an open set.
    (c) **Lower semicontinuity** is defined similarly. Work backward from the fact that the negative of a lower semicontinuous function is upper semicontinuous to give the definition in terms of lim inf's.
26. Given a compact set $K \subset \mathbb{R} \times [0, \infty)$ define

$$g(x) = \begin{cases} \max\{y : (x, y) \in K\} & \text{if } K \cap (x \times \mathbb{R}) \neq \emptyset \\ 0 & \text{if } K \cap (x \times \mathbb{R}) = \emptyset. \end{cases}$$

Prove that g is upper semicontinuous.
27. Complete the proof of Theorem 24 by removing the extra hypotheses that f is bounded and defined on a compact interval.

*28. A nonnegative linear combination of measurable characteristic functions is a **simple function**. That is,

$$\phi(x) = \sum_{i=1}^{n} c_i \chi_{E_i}(x)$$

where $E_1, \ldots, E_n$ are measurable sets and $c_1, \ldots, c_n$ are nonnegative constants. We say that $\sum c_i \chi_{E_i}$ **expresses** ϕ. If the sets E_i are disjoint and the coefficients c_i are distinct and positive then the expression for ϕ is called **canonical**.
    (a) Show that a canonical expression for a simple function exists and is unique.
    (b) It is obvious that the integral of $\phi = \sum c_i \chi_{E_i}$ (the measure of its undergraph) equals $\sum c_i m E_i$ if the expression is the canonical one. Prove carefully that this remains true for *every* expression of a simple function.
    (c) Infer from (b) that $\int \phi + \psi = \int \phi + \int \psi$ for simple functions.
    (d) Given measurable $f, g : \mathbb{R} \to [0, \infty)$, show that there exist sequences of simple functions $\phi_n \uparrow f$ and $\psi_n \uparrow g$ as $n \to \infty$.
    (e) Combine (c) and (d) to revalidate linearity of the integral.
29. Assume that $f : \mathbb{R} \to [0, \infty)$ is integrable.
    (a) Show that there exists a sequence of bi-infinite partitions Y_n of the y-axis as described in Section 5 for which the Lebesgue upper sums are finite and converge to $\int f$ as $n \to \infty$.
    (b) Find an example of an integrable function $f : \mathbb{R} \to [0, \infty)$ for which the upper Lebesgue sums do not converge to the integral as the mesh of the bi-infinite partition tends to zero. (The mesh is the supremum of the interval lengths $|y_i - y_{i-1}|$. The difference between (a) and (b) is lim inf versus lim.)
    ***(c) Is there a definition of the mesh of a bi-infinite partition of the positive y-axis such that the Lebesgue upper sums do converge to the integral as the mesh tends to 0?
30. Find a sequence of measurable functions $f_n : [0, 1] \to [0, 1]$ such that $\int f_n \to 0$ as $n \to \infty$, but for no $x \in [0, 1]$ does $f_n(x)$ converge to a limit as $n \to \infty$.
31. The **total undergraph** of $f : \mathbb{R} \to \mathbb{R}$ is the set $\{(x, y) : y < f(x)\}$. Using undergraph pictures, show that the total undergraph is measurable if and only if the positive and negative parts of f are measurable.

*32. (a) Assume that $A_n \uparrow A$ as $n \to \infty$ but do not assume that A_n is measurable. Prove that $m^*A_n \to m^*A$ as $n \to \infty$. (This is upward measure continuity for outer measure. [Hint: Regularity gives G_δ-sets $G_n \supset A_n$ with $mG_n = m^*A_n$. Can you make sure that G_n increases as $n \to \infty$? If so, what can you say about $G = \bigcup G_n$?])

(b) Is upward measure continuity true for inner measure? (Proof or counter-example.)

(c) What about downward measure continuity of inner measure? of outer measure?

**33. Prove that the outer measure of the Cartesian product of sets which are not necessarily measurable is the product of their outer measures. [Hint: Begin by assuming that A is not necessarily measurable and B is compact. Show that $A \times B$ has outer measure $m^*A \cdot mB$. It helps to note that for each $x \in A$, finitely many open rectangles in a covering of $A \times B$ suffice to cover $x \times B$.]

34. Check linearity of the integral directly for the two measurable characteristic functions, $f = \chi_F$ and $g = \chi_G$.

35. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = \begin{cases} \dfrac{1}{y^2} & \text{if } 0 < x < y < 1 \\ \dfrac{-1}{x^2} & \text{if } 0 < y < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that the iterated integrals exist and are finite (calculate them), but the double integral does not exist.

(b) Explain why (a) does not contradict Corollary 31.

36. Do (A) or (B), not both.

(A) (a) State and prove Cavalieri's Principle in dimension 4.

(b) Formulate the Fubini-Tonelli theorem for triple integrals and use (a) to prove it.

(B) (a) State Cavalieri's Principle in dimension $n + 1$.

(b) State the Fubini-Tonelli Theorem for multiple integrals and use (a) to prove it.

How short can you make your answers?

*37. Here is a trick question: "Are there any functions for which the Riemann integral converges but the Lebesgue integral diverges?" Corollary 50 would suggest the answer is "no." Show, however, that the

improper Riemann integral $\int_0^1 f(x)\,dx$ of

$$f(x) = \begin{cases} \dfrac{\pi}{x} \sin \dfrac{\pi}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

exists (and is finite) while the Lebesgue integral is infinite. [Hint: Integration by parts gives

$$\int_a^1 \frac{\pi}{x} \sin \frac{\pi}{x}\,dx = x \cos \frac{\pi}{x} \Big|_a^1 - \int_a^1 \cos \frac{\pi}{x}\,dx.$$

Why does this converge to a limit as $a \to 0^+$? To check divergence of the Lebesgue integral, consider intervals $[1/(k+1),\ 1/k]$. On such an interval the sine of π/x is everywhere positive or everywhere negative. The cosine is $+1$ at one endpoint and -1 at the other. Now use the integration by parts formula again and the fact that the harmonic series diverges.]

38. A theory of integration more general than Lebesgue's is due to A. Denjoy. Rediscovered by Henstock and Kurzweil, it is described in R. McLeod's book, *The Generalized Riemann Integral*. The definition is deceptively simple. Let $f : [a, b] \to \mathbb{R}$ be given. The **Denjoy integral of f, if it exists, is a real number I such that for each $\epsilon > 0$ there is a function $\delta : [a, b] \to (0, \infty)$ and

$$\left| \sum_{k=1}^{n} f(t_k)\Delta x_k - I \right| < \epsilon$$

for all Riemann sums with $\Delta x_k < \delta(t_k)$, $k = 1, \ldots, n$. (McLeod refers to the function δ as a **gauge** and to the intermediate points t_k as **tags**.)

(a) Verify that if we require the gauge $\delta(t)$ to be continuous then the Denjoy integral reduces to the Riemann integral.

(b) Verify that the function

$$f(x) = \begin{cases} \dfrac{1}{\sqrt{x}} & \text{if } 0 < x \leq 1 \\ 100 & \text{if } x = 0 \end{cases}$$

has Denjoy integral 2. [Hint: Construct gauges $\delta(t)$ such that $\delta(0) > 0$ but $\lim_{t \to 0^+} \delta(t) = 0$.]

(c) Generalize (b) to include all functions defined on $[a, b]$ for which the improper Riemann integral is finite.
(d) Infer from (c) and Exercise 37 that some functions are Denjoy integrable but not Lebesgue integrable.
(e) Read McLeod's book to verify that
(i) Every nonnegative Denjoy integrable function is Lebesgue integrable, and the integrals are equal.
(ii) Every Lebesgue integrable function is Denjoy integrable, and the integrals are equal.
Infer that the difference between Lebesgue and Denjoy corresponds to the difference between absolutely and conditionally convergent series: if f is Lebesgue integrable, so is $|f|$, but this is not true for Denjoy integrals.

39. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a rotation and let $E \subset \mathbb{R}^2$ be measurable. Show that $m(TE) = mE$ by completing the following outline.
(a) The planar measure of a disc is determined solely by its radius.
(b) A rectangle R is the union of a zero set and countably many disjoint discs.
(c) Under T, discs are sent to discs of equal radius, and a zero set is sent to a zero set.
(d) Infer from (a)-(c) that TR is measurable and $m(TR) = |R|$.
(e) Infer from (d) that TE is measurable and $m(TE) = mE$. [Hint: Regularity.]
(f) Generalize to $\mathbb{R}^n$.
Combined with translation invariance, this exercise shows that Lebesgue measure is invariant under all rigid motions of $\mathbb{R}^n$.

40. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is a general linear transformation, use Exercise 39 and the polar form of T explained in Appendix D of Chapter 5 to show that if $E \subset \mathbb{R}^n$ is measurable then $m(TE) = |\det T| mE$.

41. The **balanced density** of a measurable set E at x is the limit, if exists, of the concentration of E in B where B is a ball centered at x that shrinks down to x. Write $\delta_{\text{balanced}}(x, E)$ to indicate the balanced density, and if it is 1, refer to x as a **balanced density point**.
(a) Why is it immediate from the Lebesgue Density Theorem that almost every point of E is a balanced density point?
(b) Given $\alpha \in [0, 1]$, construct an example of a measurable set $E \subset \mathbb{R}$ that contains a point x with $\delta_{\text{balanced}}(x, E) = \alpha$.
(c) Given $\alpha \in [0, 1]$, construct an example of a measurable set $E \subset \mathbb{R}$ that contains a point x with $\delta(x, E) = \alpha$.

**(d) I s there a single set that contains points of both types of density for all $\alpha \in [0, 1]$?

42. Suppose that $P \subset \mathbb{R}$ has the property that for every interval $(a, b) \subset \mathbb{R}$,
$$\frac{m(P \cap (a, b))}{b - a} = \frac{1}{2}.$$
(a) Prove that P is nonmeasurable. [Hint: This is a one-liner.]
(b) Is there anything special about $1/2$?

**43. Assume that the (unbalanced) density of E exists at every point of $\mathbb{R}$, not merely at almost all of them. Prove that up to a zero set, $E = \mathbb{R}$, or $E = \emptyset$. (This is a kind of measure theoretic connectedness. Topological connectedness of $\mathbb{R}$ is useful in the proof.) Is this also true in $\mathbb{R}^n$?

44. Prove that any positive measure subset of $\mathbb{R}$ contains a positive measure Cantor set. Is the same true in $\mathbb{R}^n$?

*45. As indicated in Appendix D, $U \subset \mathbb{R}^n$ is K-quasi-round if it can be sandwiched between balls $B \subset U \subset B'$ such that $\operatorname{diam} B' \leq K \operatorname{diam} B$.
(a) Prove that in the plane, squares and equilateral triangles are (uniformly) quasi-round. (The same K works for all of them.)
(b) What about isosceles triangles?
(c) Formulate a Vitali Covering Lemma for a Vitali covering $\mathcal{V}$ of $A \subset \mathbb{R}^2$ by uniformly quasi-round sets instead of discs.
(d) Prove it.
(e) Generalize to $\mathbb{R}^n$.
(f) Consider the alternate definition of K-quasi-roundness of a measurable $V \subset \mathbb{R}^n$ as
$$\frac{\operatorname{diam}(V)^n}{mV} \leq K.$$
What is the relation between the two definitions, and is the Vitali Covering Lemma true for coverings by uniformly quasi-round sets under the second definition?
[Hint: Review the proof of the Vitali Covering Lemma.]

*46. Construct a Jordan curve (homeomorphic copy of the circle $\{(x, y) : x^2 + y^2 = 1\}$) in $\mathbb{R}^2$ that has positive planar measure. [Hint: Given a Cantor set in the plane, is there a Jordan curve that contains it? Is there a Cantor set in the plane with positive planar measure?]

47. [Speculative] Density seems to be a first order concept. To say that the density of E at x is 1 means that the concentration of E in a ball

B containing x tends to 1 as $B \downarrow x$. That is,

$$\frac{m(B) - m(E \cap B)}{mB} \to 0.$$

But how fast can we hope it tends to 0? We could call x a **double density point** if the ratio still tends to 0 when we square the denominator. Interior points of E are double density points. Are such points common or scarce in a measurable set? What about balanced density points?

48. Let $E \subset \mathbb{R}^n$ be measurable, and let x be a point of ∂E, the boundary of E. (That is, x lies in both the closure of E and the closure of E^c.)
    (a) Is it true that if the density $\delta = \delta(x, E)$ exists then $0 < \delta < 1$? Proof or counter-example.
    (b) Is it true that if $\delta = \delta(x, E)$ exists and $0 < \delta < 1$ then x lies in ∂E? Proof or counter-example.
    (c) What about balanced density?
49. Choose a pair of derivates other than the right max and left min. If f is monotone write out a proof that these derivates are equal almost everywhere.
50. Construct a monotone function $f : [0, 1] \to \mathbb{R}$ whose discontinuity set is exactly the set $\mathbb{Q} \cap [0, 1]$, or prove that such a function does not exist.

*51. Construct a strictly monotone function whose derivative is equal to zero almost everywhere.

*52. In Section 9 the **total variation** of a function $f : [a, b] \to \mathbb{R}$ is defined as the supremum of all sums $\sum_{i=1}^{n} |\Delta_i f|$ where P partitions $[a, b]$ into subintervals $[x_{i-1}, x_i]$ and $\Delta_i f = f(x_i) - f(x_{i-1})$. Assume that the total variation of f is finite (i.e., f is of **bounded variation**) and define

$$T_a^x = \sup_P \{\sum_k |\Delta_i f|\}$$

$$P_a^x = \sup_P \{\sum_k \Delta_i f : \Delta_i f \geq 0\}$$

$$N_a^x = -\inf_P \{\sum_k \Delta_i f : \Delta_i f \leq 0\}$$

where P ranges through all partitions of $[a, x]$. Prove that
    (a) f is bounded.
    (b) T_a^x, P_a^x, N_a^x are monotone nondecreasing functions of x.

    (c) $T_a^x = P_a^x + N_a^x$.
    (d) $f(x) = f(a) + P_a^x - N_a^x$.

53. Assume that $f : [a, b] \to \mathbb{R}$ has bounded variation. The **Banach indicatrix is the function

$$y \mapsto N_y = \#f^{\text{pre}}(y).$$

    (a) Prove that $N_y < \infty$ for almost every y.
    (b) Prove that $y \mapsto N_y$ is measurable.
    (c) Prove that

$$T_a^b = \int_c^d N_y \, dy$$

    where $c \leq \min f$ and $\max f \leq d$.

*54. A sequence of measurable functions $f_n : [a, b] \to \mathbb{R}$ converges **nearly uniformly** to f as $n \to \infty$ if for each $\epsilon > 0$ there is an ϵ-set $S \subset [a, b]$ (that is, $mS < \epsilon$) such that $f_n(x) \to f(x)$ uniformly as x varies in S^c and $n \to \infty$.
    (a) Contrast nearly uniform convergence with **almost uniform** convergence, which means uniform convergence on the complement of a zero set.
    (b) **Egoroff's Theorem** states that almost everywhere convergence on $[a, b]$ implies nearly uniform convergence. Prove it as follows.
        (i) Let $f_n : [a, b] \to \mathbb{R}$ be a sequence of measurable functions that converges almost everywhere to a function f, and set

$$X(k, \ell) = \{x \in [a, b] : \forall n \geq k, |f_n(x) - f(x)| < 1/\ell\}.$$

        Show that for each fixed ℓ, $\bigcup_k X(k, \ell)$ equals $[a, b]$ modulo a zero set.
        (ii) Given $\epsilon > 0$ show that there exists a sequence $k_1 < k_2 < \ldots$ such that for $X_\ell = X(k_\ell, \ell)$ we have $m(X_\ell^c) < \epsilon/2^\ell$.
        (iii) Infer that $X = \bigcap_\ell X_\ell$ has $m(X^c) < \epsilon$ and that f_n converges uniformly to f on X. (Avoid re-using the letter ϵ in your proof.)
    (c) Your proof in (b) is also valid for functions defined on any bounded subset of Euclidean space, is it not?
    (d) Why is Egoroff's Theorem false for functions defined on unbounded domains such as $\mathbb{R}$?

*55. Show that nearly uniform convergence is transitive in the following sense. Assume that f_n converges nearly uniformly to f as $n \to \infty$,

and that for each fixed n there is a sequence $f_{n,k}$ which converges nearly uniformly to f_n as $k \to \infty$. (All the functions are measurable and defined on $[a, b]$.)

(a) Show that there is a sequence $k(n) \to \infty$ as $n \to \infty$ such that $f_{n,k(n)}$ converges nearly uniformly to f as $n \to \infty$. In symbols

$$\operatorname*{nulim}_{n\to\infty} \operatorname*{nulim}_{k\to\infty} f_{n,k} = f \quad \Rightarrow \quad \operatorname*{nulim}_{n\to\infty} f_{n,k(n)} = f.$$

(b) Why does (a) remain true when almost everywhere convergence replaces nearly uniform convergence? [Hint: The answer is one word.]

(c) Is (a) true when $\mathbb{R}$ replaces $[a, b]$?

(d) Is (b) true when $\mathbb{R}$ replaces $[a, b]$?

56. Consider the continuous functions

$$f_{n,k}(x) = (\cos(\pi n! x))^k$$

for $k, n \in \mathbb{N}$ and $x \in \mathbb{R}$.

(a) Show that for each $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \lim_{k\to\infty} f_{n,k}(x) = \chi_{\mathbb{Q}}(x),$$

the characteristic function of the rationals.

(b) Infer from Exercise 23 in Chapter 3 that there can not exist a sequence $f_{n,k(n)}$ converging everywhere as $n \to \infty$.

(c) Interpret (b) to say that everywhere convergence can not replace almost everywhere convergence or nearly uniform convergence in Exercise 55.

*57. **Lusin's Theorem** states that a measurable function $f : [a, b] \to \mathbb{R}$ is **nearly uniformly continuous** in the sense that for any $\epsilon > 0$ there is an ϵ-set $S \subset [a, b]$ such that the restriction of f to S^c is uniformly continuous. Prove Lusin's Theorem as follows.

(a) Show that the characteristic function of an open interval is the nearly uniform limit of continuous functions.

(b) Infer from (a) that the same is true of the characteristic function of a measurable set. [Hint: Regularity and Exercise 55.]

(c) Infer from (b) that the same is true for a simple function.

(d) Use Egoroff's Theorem and Exercises 28, 55 to infer from (c) that the same is true for a nonnegative measurable function: it is the nearly uniform limit of a sequence of continuous functions.

(e) Given a measurable $f : [a, b] \to \mathbb{R}$ and given $\epsilon > 0$ infer from (d) that there exists a sequence of continuous functions $f_n : [a, b] \to \mathbb{R}$ and an open ϵ-set $U \subset [a, b]$ such that $f_n \to f$ uniformly on U^c as $n \to \infty$.

(f) Why does (e) imply that f is nearly uniformly continuous?

58. At what stage, if any, does your reasoning in Exercise 57 make essential use of one-dimensionality? Explain.

59. Let $f : \mathbb{R} \to \mathbb{R}$ be measurable.

(a) Give an example of an f for which the conclusion of Lusin's Theorem is false.

(b) Formulate the definition of **nearly continuous** and prove that f is nearly continuous.

(c) Generalize to $\mathbb{R}^n$.

*60. Let E be a measurable subset of the line having positive Lebesque measure.

(a) Prove Steinhaus' Theorem: E meets its t-translate for all sufficiently small t. [Hint: density points.]

(b) Formulate and prove the corresponding result in higher dimensions.

(c) Prove that despite the fact that the standard Cantor set has measure zero, it meets each of its t-translates for $|t| \leq 1$.

Index

1-form, 314
C^r equivalent, 291
C^r norm, 285
F_σ-set, 189
G_δ-set, 189
M-test, 207, 211, 241, 285
α-Hölder, 186, 253
δ-dense, 253
ϵ-chain, 123
ϵ-principle, 21
ϵ, δ-condition, 55
$\epsilon/2^n$-argument, 164
σ-compact, 250, 257
k-cell, 315
k-chain, 325
k-form, 315
p-adic, 130
p-series, 182
r^{th} derivative, 147
t-advance map, 234
x_I-area, 315

absolute continuity, 398
absolute convergence of a series, 181
absolute property, 82
abuse of notation, 7
accumulation points, 69
address string, 97
adheres, 59
advance map, 234
aleph null, 30
algebraic number, 48
almost everywhere, 163
almost uniform convergence, 427
alternating harmonic series, 184
alternating series, 184
ambient homeomorphism, 105
ambiently diffeomorphic, 357
analogy, 10
analytic function, 148, 235
Analyticity Theorem, 237
antiderivative, 173
Antiderivative Theorem, 173
Antoine's Necklace, 107
arc, 122
Archimedean property, 20
area, 306, 364
argument by contradiction, 8
Arzela-Ascoli Theorems, 214, 216
ascending k-tuple index, 317
ascending presentation, 317
associativity, 14
average (of a function), 396
average derivative, 278

Baire class 1, 189
Baire's Theorem, 243
baker's transformation, 118
Banach Contraction Principle, 228
Banach indicatrix, 427
Banach space, 285
basic k-form, 315
Bernstein polynomial, 218
bijection, 30
bilinear, 275
block test, 197
Bolzano-Weierstrass Theorem, 77
Borel's Lemma, 256
boundary of a k-cell, 325
boundary of a set, 66, 127
bounded function, 249
bounded linear transformation, 268
bounded metric, 124
bounded sequence, 49
bounded variation, 406, 426
box, 24
Brouwer's Fixed Point Theorem, 228, 334
bump function, 188
Burkill, J.C., 376

Cantor function, 175
Cantor piece, 103
Cantor set, fat, 98, 193
Cantor set, middle quarters, 192
Cantor set, middle-thirds, 95
Cantor space, 103
Cantor surjection theorem, 99
cardinality, 28
Cartesian product, 21
Cauchy completion, 112
Cauchy condition, 18, 73
Cauchy Convergence Criteria, 19, 180
Cauchy product of series, 199
Cauchy sequence in a metric space, 73
Cauchy-Binet Formula, 323, 343
Cauchy-Riemann Equations, 340
Cauchy-Schwarz Inequality, 22
Cavalieri's Principle, 305, 338
center of a starlike set, 121
chain connected, 252
chain rule, 274
chain-connected, 123
Change of Variables Formula, 306
characteristic function, 160
Chebyshev Lemma, 401
class, 2
class C^0, C^1, C^r, C^∞, C^ω, 147
clopen, 60
closed form, 329
closed neighborhood, 94
closed set, 59
closed set condition, 65
closure of a set, 66
cluster point, 69, 126
co-Cauchy, 109
coarse partition, 155
cohomology classes, 333
common refinement, 158
commutative diagram, 291
compact (in the covering sense), 88
compact (sequentially), 76
comparability of metrics, 71
comparable norms, 346
comparison test, 180
complement of a set, 41
complete, 13, 18
completeness of a metric space, 74
Completion Theorem, 108
complex analytic, 238
complex derivative, 339
complex differentiable, 239
complex linear map, 339
composite, 30
compound word (as an address), 100
concentration, 395
condensation point, 69, 126
condition number, 341
conditional convergence of a series, 181
cone over a set, 125
connected, 83
connected component, 136
conorm, 345
constant rank, 290
content, 419
continuity, 36, 55, 65
continuously differentiable, 147
Continuum Hypothesis, 30
contraction, 228
convergence of a sequence, 17, 54
convergence with respect to an order relation, 191
convex, 25
convex combination, 26, 45
convex function, 46
convex hull, 105
countable, 30
countable additivity, 364, 368
countable base of a topology, 128
covering, 88
covering compact, 88
covering, Vitali or fine, 392
critical point, 193
critical value, 193
cupcake theorem, 134
curl of a vector field, 329
cut, 11

Darboux continuous, 144
Darboux integral, 156
de Rham cohomology group, 333
decimal expansion, 42
decreasing sequence, 117
Dedekind cuts, 11
deMorgan's Law, 41
Denjoy integral, 423
dense, 98
density, 395
density point, 395
density, balanced, 395, 424
denumerable, 30
derivate, 403
derivative, 139, 271
derivative growth rate, 235
determinant, 343
Devil's ski slope, 177
Devil's staircase function, 175
diagonal, 43
diameter, 79
diffeomorphism, 152, 289
difference of sets, 2
differentiable, 141, 271
differential forms, 313
dipole, 326
directional derivative, 348
disconnected, 83
discontinuity of the first kind, 194
discontinuity of the second kind, 194
discrete metric, 52
disjoint sets, 2
distance from a point to a set, 115
distance vector, 124
divergence of a series, 179
divergence of a vector field, 328
domain of a function, 28
domination, 180
dot product, 22
double density point, 426
double integral, 390
dyadic, 387
dyadic filtration lemma, 100
dyadic number, 44
dyadic ruler function, 193

efficient covering, 392
Egoroff's Theorem, 427
embeds, 81
empty set, 2
engulf, 257
envelopes, upper and lower, 380
equal cardinality, 30
equicontinuity, 213
equivalence relation, classes, 3
error factor, 275
Euclidean distance, 24
Euler characteristic, 45
Euler's product formula, 200
exact form, 329
excludes, 94
exponential growth rate, 183
extends to (of a function), 118
exterior derivative, 321

F sigma set, 373
fat Cantor set, 98, 193
Fatou's Lemma, 382
field, 15
fine partition, 155
finite additivity, 369
finite cardinality, 30
finite intersection property, 120
finite subadditivity, 419
finite word, 48
first orthant, 24
fixed point, 43, 228, 334
flow, 233
flux of a vector field, 329
Fréchet derivative, 272

front inclusion k-cell, 326
Fubini's Theorem, 304
function, 28
function algebra, 223
functional, 314
Fundamental Theorem of Calculus, 171
Fundamental Theorem of Continuous Functions, 39

G delta set, 373
gap interval, 103
gauge, 423
general linear group, 351
generic, 243
geometric series, 180
gradient vector field, 298
grand intersection, 120
graph, 43
grid, 300
growing steeple, 204

Hahn-Mazurkiewicz Theorem, 132
Hairy Ball Theorem, 359
harmonic series, 180, 184
Hausdorff metric, 133, 199, 252
Hawaiian earring, 53, 122
Heine-Borel Theorem, 77
Heine-Borel Theorem in $C^0([a,b],\mathbb{R})$, 217
Henstock, 423
higher order differentiability, 147
Hilbert cube, 131
Hölder condition, 186, 253
holomorphic, 239
homeomorphism, 57
hull, 374, 417

identity map, 30
Identity Theorem, 256
image set of a function, 28
implicit function, 286
Implicit Function Theorem, 286
improper Riemann integral, 178
inclusion cell, 318
increasing sequence, 117
indefinite Lebesgue integral, 398
indicator function, 160
infimum, 17
infinite (cardinality), 30
infinite product, 198
infinitely differentiable, 147
inheritance, 52, 67
Inheritance Principle, 68
initial condition, 230
injection, 29
inner measure, 364, 374
inner product, 27
inner product space, 27
integer lattice, 24
integers, 1
integral test, 181
integrally equivalent functions, 194
integration by parts, 178
integration by substitution, 177
integration operators, 330
interior of a set, 66, 127
intermediate value property, 38, 144
Intermediate Value Theorem, 38, 83, 84
interval, 19
intrinsic property, 82
Inverse Function Theorem, 152, 289
inverse image, 64
irrational numbers, 19
isometry, 116
iterate (of a function), 407
iterated integral, 390

Jacobian, 306
Jordan content, 306, 419
Jordan Curve Theorem, 132
jump discontinuity, 46, 194

kernel, 374, 418
Kurzweil, 423

L'Hospital's Rule, 143
Lagrange multiplier, 296
lattice, 24
least upper bound property, 13
Lebesgue Density Theorem, 395
Lebesgue Dominated Convergence Theorem, 381
Lebesgue integrable, 377
Lebesgue integral, 376
Lebesgue measure, 367
Lebesgue number, 89, 166

Lebesgue outer measure, 363
Lebesgue sums, 383
Lebesgue's Fundamental Theorem of Calculus, 396
Lebesgue's Last Therorem, 401
Lebesgue's Main Theorem, 399
left limit, 46
Leibniz rule, 274
length, 363
length of a vector, 22
less than or equal, 13
letters in an address, 100
limit, 17, 54, 58
limit infimum, 49
limit point, 58
limit set, 61
limit supremum, 49
linear transformation, 267
Lipschitz condition, 142, 170, 195, 230
Littlewood's princinples, 394
locally path-connected, 132
logarithm function, 174
lower bound, 43
lower integral, 156
lower semicontinuity, 420
lower sum, 156
Lusin's Theorem, 428

magnitude, 16
magnitude of a vector, 22
Manhattan metric, 72, 124
maximum stretch, 268
McLeod, 423
meager, 243
mean value property, 141
Mean Value Theorem, 141, 277, 278
measurability (of a function), 376
measurability, pre-image, 383
measurable, 367
measurable set, 367
measure continuity, 370
Mertens' Theorem, 199
mesh of a partition, 155
metaphor, 9
metric, metric space, 51
middle quarters Cantor set, 192
middle-thirds Cantor set, 95
minimum stretch, 345
modulus of continuity, 253
monotone, 117
Monotone Convergence Theorem, 377
monotone function, 47
monotone sequence, 117
monotonicity of outer measure, 364
Moore-Kline Theorem, 103
most (in the Baire sense), 243

name of a form, 314
natural numbers, 1
nearly, 392, 412
nearly continuous, 412, 429
nearly uniform continuity, 428
nearly uniform convergence, 412, 427
neighborhood, 62
nested sequence, 78
net x-variation, 314
norm, 27
normed space, 268
nowhere dense, 98
nowhere differentiable functions, 240

one-to-one, 29
onto, 29
open covering, 88
open map, 118
open set condition, 65
operator norm, 268
orbit, 118, 407
ordered field, 15
ordinate set, 377
orthant, 24
orthogonal linear map, 341
oscillating discontinuity, 194
oscillation, 165
outer measure, 363
outer measure axioms, 364
outer measure, abstract, 367

parallelogram law, 50
partial derivative, 272
partition, 104
partition of an index set, 353
partition pair, 154
path, 86
path-connected, 86, 132
Peano curve, 102
Peano space, 132

perfect, 93
periodic, 407
permutation, 342
Picard's Theorem, 231
piece, 99
piecewise continuous function, 160
piecewise linear function, 245
planar outer measure, 364
Poincaré's Lemma, 330
pointwise convergence, 201
pointwise equicontinuity, 213, 249
pointwise limit, 201
polar form, 341
positive definite symmetric linear map, 341
positive definiteness of a metric, 52
power series, 185, 211
power set, 47
pre-image, 64
prime number, 6
product matrix, 270
propagation, 216
proper subset, 82
pullback, 322
pushforward, 322

quasi-round, 412

Rademacher's Theorem, 406
radius of convergence, 185
range of a function, 28
rank, 290
Rank Theorem, 292
Ratio Mean Value Theorem, 143
ratio test, 183
rational cut, 12
rational numbers, 2
rational ruler function, 161
real number, 12
rear inclusion k-cell, 326
rearrangement of a sequence, 118
rearrangement of a series, 198
reduces to (of a covering), 88
refinement of a partition, 158
Refinement Principle, 158
regularity (of Lebesgue measure), 373
regularity hierarchy, 148, 208
relative measure, 395
representative of an equivalence class, 3
residual, 243
retraction, 334
Riemann ζ- function, 200
Riemann integrable, 155, 300
Riemann integral, 155
Riemann measurable, 306
Riemann sum, 154, 300
Riemann's Integrability Criterion, 160
Riemann-Lebesgue Theorem, 163, 302
right interval, 403
right limit, 46
right slope, 403
root test, 183

sawtooth function, 240
Schroeder-Bernstein Theorem, 34
scraps, 88
second derivative, 280
second-differentiable, 280
semicontinuity, 420
separable metric space, 128
separates points, 223
separation (disconnectedness), 83
sequence, 17, 52
sequentially compact, 76
sigma algebra, 368
sign of a permutation, 342
signed commutativity, 316
simple k-form, 315
simple function, 386, 421
simple region, 356
simply connected, 330
singleton set, 2
slice, 387
sliding secant method, 145
slope over an interval, 401
smooth, 147, 284
somewhere dense, 98
space filling, 102
staircase approximation, 356
starlike, 121, 332
steeple functions, 204
Steinhaus' Theorem, 429
step function, 160
Stokes' Theorem, 326, 327
Stone-Weierstrass Theorem, 223
strictly decreasing sequence, 117

strictly increasing sequence, 117
subadditivity, finite, 419
subcovering, 88
subfield, 15
sublinear, 271
subsequence, 54
subspace, 52
sup norm, 205
supersolving, 225
supremum, 17
surjection, 29
symmetric difference, 2
symmetrization of a bilinear map, 350
symmetry of a metric, 52

tag, 423
tail of a series, 180
tame, 106
target of a function, 28
taxicab metric, 72, 124
Taylor polynomial, 149
Taylor series, 151, 235
Taylor's Theorem, 150, 238
telescoping sum, 207
term by term differentiation, 210
term by term integration, 208
thick, 243
thin, 243
tiling a manifold with cells, 327
topological equivalence, 66
topological property, 64
topological space, 60
topologist's sine circle, 53, 123
topologist's sine curve, 87
topology, 60
total derivative, 272
total length, 98, 363
total length of a covering by intervals, 163
total variation, 406, 426
totally bounded, 92
totally disconnected, 96
trajectory, 230
transitive relation, 191
transitivity, 15
translation, 15
translation (by a function), 378
trefoil knot, 105, 134
Triangle Inequality, 16
triangle inequality, 23, 24, 52
trichotomy, 15
trigonometric polynomial, 227
truncate an address string, 97
type 1 region, type 2 region, 356

ultrametric, 129
uncountable, 30
undergraph, 154, 376
undergraph, completed, 380
undergraph, total, 382, 421
uniform C^r convergence, 284
uniform continuity, 48, 82, 118
uniform convergence, 201
uniform equicontinuity, 249
uniform limit, 202
uniformly C^r Cauchy, 284
unit ball, cube, sphere, 24
universal compact metric space, 99
upper bound, 13
upper integral, 156
upper semicontinuity, 136, 264, 420
upper sum, 156
utility problem, 132

vanishing at a point, 223
variation, 406
Vitali covering, 392
volume multiplier, 306

weak contraction, 254
wedge product, 318
Weierstrass M-test, 207
Weierstrass Approximation Theorem, 217
weighting signed area, 315
wild, 106
word (as an address), 100

Zeno's maze, 53
Zeno's staircase function, 161
zero locus, 256, 286
zero set, 98, 163, 302, 365
zero-th derivative, 147